Aalto University
School of Science
Degree Programme in Computational and Systems Biology (euSYSBIO)

Yogesh Dhungana

Truncated small chaperone HSPB1 in peripheral neuropathy:

molecular mechanisms and altered cellular pathways

Master's Thesis
Espoo, July 31, 2015

Supervisor: Professor Juho Rousu, Aalto University
Professor Erik Aurell, KTH Royal Institute of Technology
Advisor: Henna Tyynismaa Docent, University of Helsinki
Elena Czeizler PhD, Aalto University

Aalto University
School of Science
Degree Programme in Computational and Systems Biology (euSYSBIO)

ABSTRACT OF MASTER'S THESIS

Author: Yogesh Dhungana

Title:
Truncated small chaperone HSPB1 in peripheral neuropathy: molecular mechanisms and altered cellular pathways

Date: July 31, 2015 **Pages:** 58

Major: Computational Systems Biology **Code:** T-61

Supervisor: Professor Juho Rousu
Professor Erik Aurell

Advisor: Henna Tyynismaa Docent
Elena Czeizler PhD

HSPB1 is a member of a family; small heat shock proteins (sHSP). These sHSP have a common conserved α-crystallin domain (ACD), flanked by variable N- and C- termini, whose functions are poorly understood. More than 20 disease causing mutations in sHSPs has been described till date, most of them being dominant missense mutations clustered around the ACD. The protein is ubiquitously expressed, participates in a number of cellular processes, and different pathogenic mechanisms have been proposed. In this study, we investigated the cellular, transcriptional and biochemical consequences of a novel heterozygous frame shift mutation in *HSPB1* p.(M169Cfs2X) predicting C-terminal truncation of the protein in a patient with dominantly inherited motor predominant axonal Charcot-Marie-Tooth disease. We show that the truncated protein is stable in primary fibroblasts, proving that the mutant mRNA is able to escape nonsense mediated decay. Moreover, the truncated protein binds wild type *HSPB1* and interferes with its dimerization. The mutant protein translocates to nucleus and impairs the heat tolerance of patient primary fibroblasts. However, the mutant *HSPB1* does not alter the global transcriptional response to heat. We suggest that ablation of the *HSPB1* C-terminus might have caused neuropathy by a dominant negative mechanism that prevents normal *HSPB1* dimerization. The impairment of the cellular growth is mediated by the inability of truncated *HSPB1* to chaperone proteins under heat stress.

Keywords: CMT, dHMN, Heat Shock Protein, HSPB1, Protein misfolding, nonsense-mediated decay, dominant negative effect

Language: English

Acknowledgements

First and foremost, I would like to express my gratitude towards euSYSBIO consortium and Prof. Erik Aurell, Prof. Juho Rousu and Prof. Isabel Sá Correia; coordinators of the program for providing me an opportunity and scholarship to pursue my master's degree in Computational and Systems Biology. All professors and fellow students are thanked who supported and helped me in my pursuit.

I am thankful to Prof. Juho Rousu and Prof. Erik Aurell for being my thesis supervisors.

I would like to express my sincere gratitude towards Henna Tyynismmaa for giving me an opportunity to work on her group and guiding me through the project. Your support and guidance since last summer has been really rewarding and you have been truly inspirational person.

I am grateful to Elena Czeizler for her support in computational aspect of the thesis and for organizing wonderful courses; High-throughput Bioinformatics and Computational Genomics. These courses taught me statistics and computational methods, which were part of my thesis work.

I would like to thank Emil Ylikallio for his support and help for the experimental part of the project. It was a privilege to work on your project.

I warmly thank all members of Tyynismmaa group; Svetlana, Taru, Maria and Ritta for helping me understand the background and teaching me standard lab practices and making my stay wonderful and enjoyable.

Espoo, July 31, 2015

Yogesh Dhungana

Abbreviations and Acronyms

ACD	α-crystalline domain
BSA	Bovine serum albumin
CHCHD10	Coiled-coil-helix-coiled-coil-helix-domain containing protein 10
CMT	Charcot-Marie-Tooth
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DAPI	4',6-diamidino-2-phenylindole
dHMN	distal hereditary motor neuropathy
DMEM	Dulbecco's modified Eagle's medium
DNA	Deoxyribonucleic acid
DNAJB1	DnaJ(Hsp40) homolog, subfamilyB, member 1
EGR1	Early Growth Response 1
EGTA	Ethylene glycol tetraacetic acid
FOS	FBJ murine osteosarcoma viral oncogene homolog
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GDAP1	Ganglioside-induced Differentiation-associated Protein 1
GJB1	Gap junction beta-1 protein
GO	Gene Ontology
HCl	Hydrochloric acid
HSPA6	Heat Shock 70 kDa Protein 6
HSPA7	Heat Shock 70 kDa Protein 7
HSPB	Heat Shock Protein Beta
HSPB1$^{\triangle \text{C-term}}$	C-terminal truncated HSPB1
HSPB1$^{-/-}$	HSPB1 knockout
KH_2PO_4	Potassium phosphate monobasic
KLF	Kruppel-like Factors
MCVs	Motor nerve conduction velocities
MFN2	Mitofusin 2

MOPS	3-(N-morpholino)propanesulphonic acid
MPZ	Myelin Protein Zero
MTMR2	Myotubularin Related Protein 2
NCBI	National Center for Biotechnology Information
NaCl	Sodium chloride
NaF	Sodium fluoride
Na_3VO_4	Sodium orthovanadate
NDRG1	N-myc Downstream Regulated Gene 1 Protein
NEF2	Nucleotide Excision Repair Factor 2
NIH	National Institute of Health
Pen-Strep	Penicillin Streptomycin cocktail
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline with Tween 20
PMP22	Peripheral Myelin Protein 22
PRX1	Paired-Related Homeobox Gene
RIPA	Radioimmunoprecipitation assay buffer
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
SBF2	SET Binding Factor 2
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
sHSP	small heat shock protein
TBST	Tris-buffered saline with Tween 20
TSS	Transcription start site
$ZnCl_2$	Zinc chloride

Contents

List of Original Publication

Partial work of the thesis is included in the following publication:

Ylikallio, E., Konovalova, S., Dhungana, Y., Hilander, T., Junna, N., Partanen, J. V., and Tyynismaa, H. (2015). Truncated HSPB1 causes axonal neuropathy and impairs tolerance to unfolded protein stress. BBA Clinical, 3, 233-242.

Chapter 1

Introduction

1.1 What is Charcot-Marie-Tooth (CMT) disease?

CMT is a group of hereditary disorders that causes damage to peripheral nerves, the nerves that transmit signals from brain and spinal cord to and from the distal part of the body, including sensory information like touch and pain [58]. CMT has a direct effect on nerves that control the muscles [48]. The progressive muscle weakness may become noticeable in adolescence or early adulthood. Clinical symptoms of CMT include length dependent degeneration of peripheral nerves that result in weakness and waste of distal muscles in extremities of body [3]. Symptoms normally begin in lower limbs and moves toward the body [42]. Majority of population with CMT have some kind of minor physical disability. According to the article published by the US Department of Health and Human Services, NIH, on Charcot-Marie-Tooth Disease, CMT is one of the most common inherited neurological disorders affecting estimated 126,000 individuals in the United States and 2.6 million people worldwide [42].

1.2 General classification of CMT

Although there is no universal classification for CMT, on the basis of upper limb motor nerve conduction velocities (MCV), we can distinguish 2 types of CMT: demyelinating and axonal types [34, 48]. Patient with MCVs below 38 m/s are classified as CMT1 whereas patients with upper limb MCV normal or slightly above 38 m/s are assigned as CMT2 cases [48]. The generic classification of CMT phenotype based on nerve conduction velocities is shown in

the Figure 1.1. Furthermore, CMT2 patients show wider range of phenotypes than CMT1 patients. This makes clinical diagnosis of CMT2 from CMT1 inconclusive until MCV is performed.

Mutations in different genes produce similar phenotypes, which leads to clinical heterogeneity in patients with CMT [3]. Until now, mutations in more than 40 genes have been associated with CMT neuropathy and related disorders [3]. The varying degree of clinical phenotypes in CMT is categorized by age dependent penetrance. The onset of disease can take place as early as a first decade of life depending on the inheritance of CMT, which can be of three types: autosomal dominant, autosomal recessive or X-linked [48].

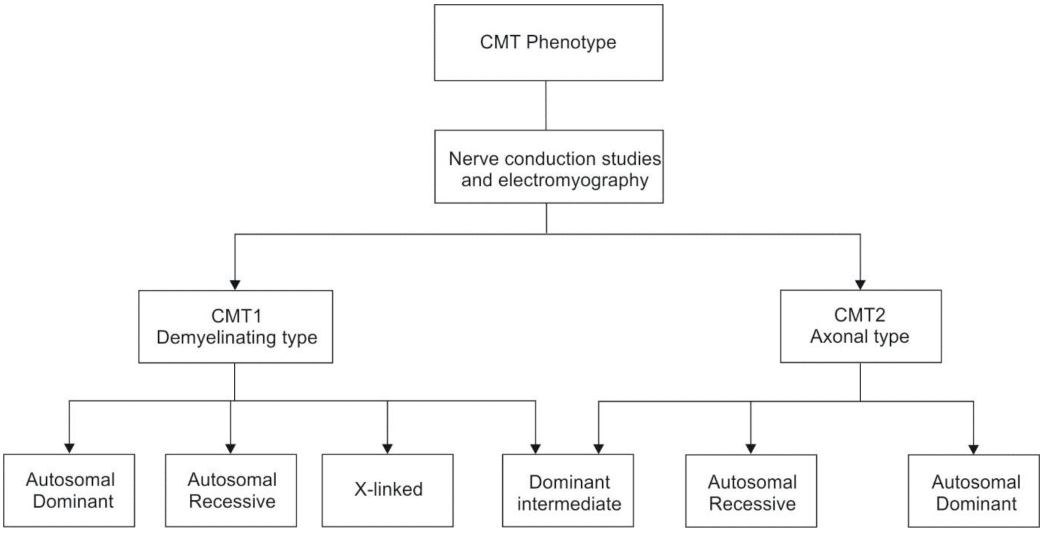


Figure 1.1: Classification of CMT phenotypes based on neurophysiology. Autosomal dominant intermediate forms of CMT can share features of both CMT1 and CMT2 phenotypes. Reproduced from [34] for the purpose of this discussion.

1.3 Different types of genetic inheritance

Diseases caused by mutation in single gene follow simple Mendelian or non-Mendelian inheritance pattern, which are divided into six different types:

1.3.1 Autosomal Dominant

The presence of one copy of mutant alleles in cell is enough to express the autosomal dominant condition in individuals. The mutation is located on one of the 22 pairs of autosomes (non-sex determining chromosomes). The

affected individual carries a copy of normal and a copy of the mutant gene inherited usually from affected parent. Huntington disease has an autosomal dominant inheritance pattern [43].

1.3.2 Autosomal Recessive

When individuals have two copies of the mutant allele, they express autosomal recessive condition. However, the presence of a single copy of the mutant allele is not enough to develop the condition but are classified as "carrier" for the mutation. Sickle cell anemia is an example of such inheritance [43].

1.3.3 X-linked Recessive

X-linked conditions are the result of mutations in genes on the X chromosome (sex determining chromosome). X-linked recessive condition develop when there is no normal copy of the gene in the X chromosomes. Thus, males have more chances to develop X-linked recessive condition as they have only one copy of the X chromosome. X-linked inheritance does not have male to male transmission and all daughters of affected male are carriers (heterozygous) of X-linked recessive condition. Hemophilia is an example of X-linked recessive disorder [43].

1.3.4 X-linked Dominant

X-linked dominant conditions are observed when there is a mutant copy of the gene in one of the X-chromosome. Thus, all daughters of affected male are affected since males have only one copy of X chromosome [43].

1.3.5 Codominant inheritance

It is the inheritance pattern observed in the ABO blood group system where each allele influences the trait. The combined effect of both allele results in characteristic genetic condition [43].

1.3.6 Mitochondrial inheritance

Mitochondria are inherited in offsprings only from the mother, so it is also called maternal inheritance and is only applied to mitochondrial genes. Mitochondrial disorders arising from mutations in mitochondrial genes can affect both male and female members of the family [43].

1.4 Classes of mutation

Mutations are heritable alteration in the sequence of the genome of an organism. The organism carrying one or more such mutations is referred to as mutant. A simple method of classifying mutations is whether or not they are point mutations, which are the result of single base pair substitution and insertion or deletion of few base pairs in the DNA sequence [26].

1.4.1 Base substitution mutation

A point mutation caused by substitution of one base pair with another is known as base substitution mutation. It can be substitution of one purine with another purine or one pyrimidine with another; commonly referred to as *transition mutation*. On the other hand, the exchange of a purine for a pyrimidine or vice versa is referred to as *transversion mutation* [26]. Furthermore, these base substitution mutations are subclassified on the basis of its effect on protein-coding regions of the genome.

1.4.1.1 Missense mutation

Base substitution in this type of mutation changes the codon of one amino acid to another. Such mutation can result in mutant protein with: no detectable effect, partial loss of function, gain of function, alteration of function, altered stability or complete loss of function. Apart from these noted effects, there are types of missense mutations referred to as *silent* or *neutral* mutations, which is the result of the degeneracy of the genetic code. These silent mutations are result of base substitution mutations caused by changing codon for an amino acid to a different codon for the same amino acid ultimately resulting in unchanged protein sequence [26].

1.4.1.2 Nonsense mutation

Mutation where the codon for an amino acid is changed to one of the three stop codons: TAG, TAA or TGA, leading to premature termination of protein synthesis is referred as nonsense mutation. These nonsense mutations can result in partial or complete loss of protein function [26].

1.4.2 Insertion or Deletion mutation

This category of mutation is a result of insertion or deletion of a few nucleotides from the DNA sequences. Usually insertion or deletion of 1, 2, 4 or

5 nucleotides occurs in case of this type of mutation and shifts the reading frame of the protein coding region of a gene, hence also termed as *frameshift mutation*.

1.5 Genetic heterogeneity of CMT

Most of the CMT2 cases have an autosomal dominant inheritance and late onset of the disease. However, rare and recessively inherited cases can show an early onset [51]. Similarly, the inheritance pattern of sporadic cases is more difficult to determine. A common pathological marker of the peripheral neuropathy is chronic axonal degeneration and regeneration, which leads to a steady loss of available nerve fibers [60]. The associated genes are known to be involved in various pathways such as myelination, radial and axonal transport, Schwann cells differentiation, signal transduction, mitochondrial function, endosome, protein translation and single stranded DNA break repair [55]. Similarly, another important pathway involved in axon maintenance is the cellular stress response pathway mediated by small heat shock proteins (sHSPs) [58].

CMT, a motor and sensory neuropathy, is closely related to other inherited neuropathies such as distal hereditary motor neuropathies (dHMN), which involves only motor nerves and hereditary sensory neuropathies (HSN), which involves only sensory nerves. The collective term used to refer all three diseases is 'CMT and related disorders' [48]. The genes implicated in these diseases include structural proteins like *PMP22* and *MPZ* associated with myelination, *GJB1* and *NEFL* genes involved in radial and axonal transport, genes of signal transduction pathways (*PRX, MTMR2, SBF2, NDRG1*), proteins related to mitochondrial function (*MFN2, GDAP1*) and molecular chaperones (*HSPB1, HSPB8*) [55]. Most recent findings include a *CHCHD10* variant p.Gly66Val in patients with CMT2 phenotype [5], p.Gly58Arg variant of the same gene causing a pure myopathy [2] and p.Ser59Leu causing more aggressive amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) [8]. An illustration of this heterogeneity is represented in Figure 1.2.

1.6 HSPB1 and Charcot-Marie-Tooth Disease

HSPB1 belongs to a group of small heat shock proteins (sHSPs) assumed to have a crucial role of protection against heat or unfolded protein stress [58]. sHSPs are ubiquitously expressed molecular chaperones with a characteris-

tic conserved α -crystallin domain bordered by variable N- and C- terminal regions [3]. Contrary to heat shock proteins with ATPase domain, sHSPs cannot actively refold unfolded proteins [3]. However, they are capable of binding unfolded proteins and keeping them in a folding competent state until acted upon by ATP-dependent HSPs [3]. These sHSPs are known to form complex oligomeric structures considered crucial for protein function. The formation of these complex structures depend upon their phosphorylation state and dimerization through β7 strands of ACD [9, 28]. Among the 10 sHSPs encoded by the human genome, mutations in three (encoded by genes *HSPB1, HSPB3* and *HSPB8*) have been reported to cause CMT2 or dHMN [25, 30, 32]. Furthermore, HSPB5 variants are known to cause myopathy and cardiomyopathy [57].

HSPB1 is a 205 amino acid protein where ACD is located at residues 87-168 [58]. The structure of *HSPB1* with reported mutations is shown in Figure 1.3. Among the functions of *HSPB1*, protection of cell against stress is thought to be the most important one [58]. Studies have demonstrated the ability of *HSPB1* in preventing in vitro aggregation of misfolded proteins [54, 58]. Nuclear translocation of *HSPB1* upon heat stress has been observed in different cell types, which also suggests a role of *HSPB1* as a chaperone of nuclear proteins or in regulation of gene expression [14, 58].

There are about 20 disease causing mutations reported in *HSPB1* and all mutations are associated with motor neuropathy [58]. Although most of the reported dominantly inherited mutations are clustered in ACD region, few reported missense mutations are in non-conserved residues of the variable N- and C- terminus of the protein [12]. It is not yet completely understood why *HSPB1* mutations lead to a disease specific to peripheral neurons despite the gene being ubiquitously expressed. Several studies have been published where different *HSPB1* mutations have been studied through overexpression experiments. Among them, missense mutations in the ACD region of *HSPB1* have been associated with defective dimerization, increased chaperone activity and improved heat tolerance [4, 58]. Studies involving mutations HSPB1^{S135F} and HSPB1^{P182L} discovered aggregation and altered axonal transport of neurofilament [1, 25, 58, 59]. These mutants had increased capacity to bind tubulin resulting in altered dynamics of microtubules [3]. In addition to this, transgenic expression of these mutants in mouse neurons demonstrated decreased acetylated α-tubulin and induced defects in axonal transport [23]. Furthermore, missense mutations in the C-terminus of *HSPB1* such as HSPB1^{P182S} and HSPB1^{R188W} were associated with aggregation and decreased chaperone activity respectively [16]. These studies provide us with varied in vitro effects

of C-terminal variants.

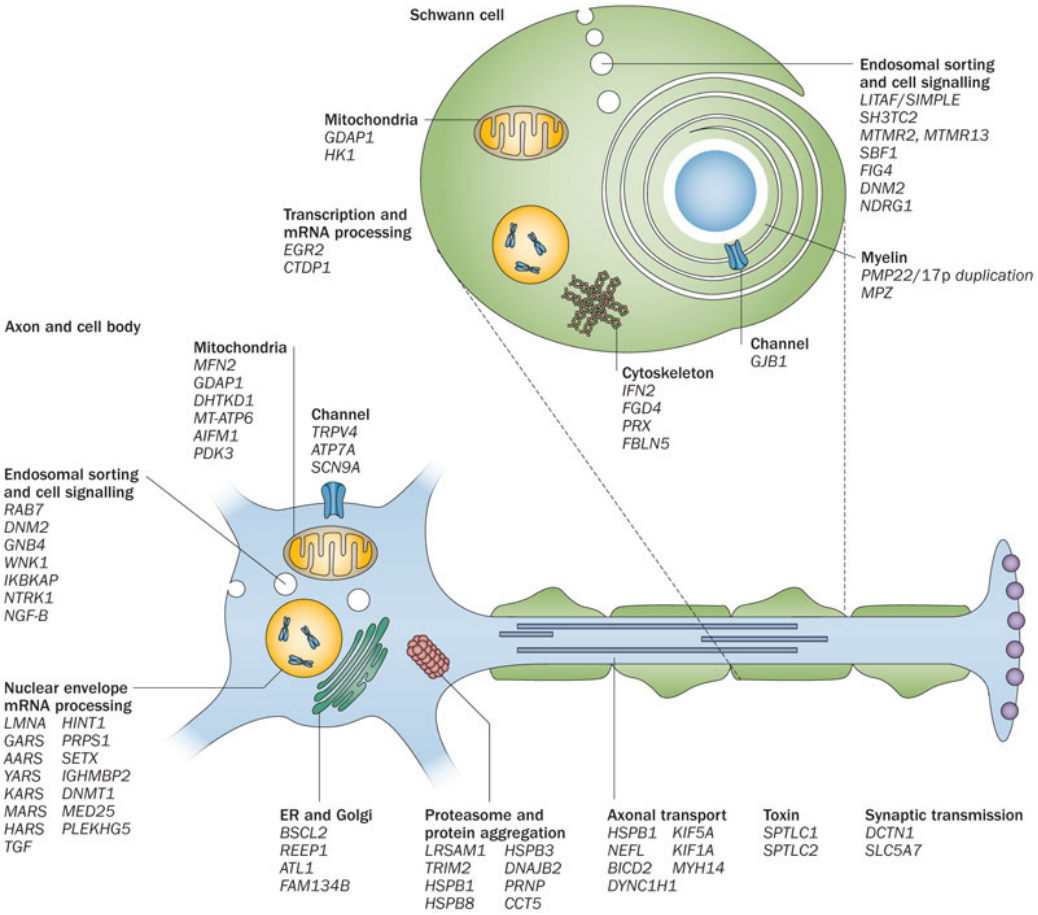


Figure 1.2: Known disease genes for CMT and related disorders, and their proposed pathomechanism. Reproduced from [51] for the purpose of this discussion.

1.7 Brief overview of Nonsense-mediated decay and dominant negative mechanism

Nonsense-mediated mRNA decay (NMD) is a highly conserved quality control mechanism present in all eukaryotes [17]. It selectively degrades mRNAs that have premature termination codons (nonsense codons) which, if translated, can produce a truncated protein with dominant-negative or potential deleterious effects [17].

Premature termination codons can arise in different ways, which includes random nonsense or frameshift mutations in the coding region of the DNA

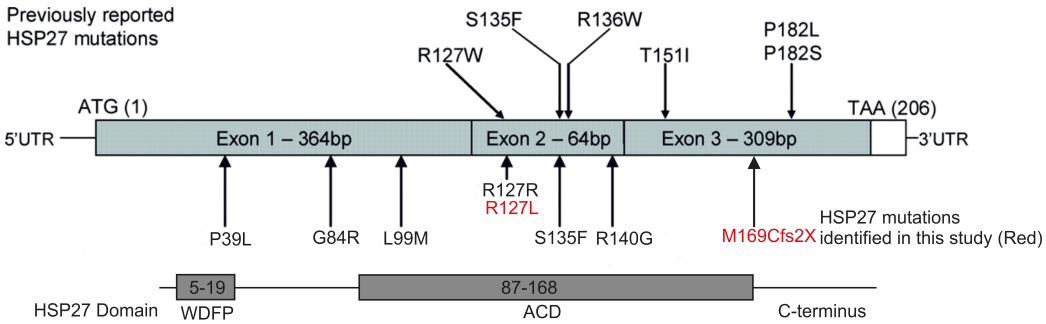


Figure 1.3: *HSPB1* also known as HSP27 contains a hydrophobic N-terminal domain with WDFP motif, an α-crystallin domain at residues Glu87-Pro168 and variable C-terminal domain from 169-205. Adapted from [29] for the purpose of this discussion.

sequences and errors in RNA splicing [17]. Studies on the mechanism of mammalian NMD suggest that it occurs only on newly synthesized Cap-binding complex bound (CBC) mRNAs [17]. Mammalian NMD also involves Exon-Junction Complex (EJC) deposited at exon-exon junctions during mRNA splicing. The general consensus about mammalian NMD being able to distinguish between premature and normal stop codon is the presence of the second signal downstream of the stop codon. The second signal in mammalian cells is different compared to yeast as it is delivered by an intron downstream of the stop codon and that intron must be spliceable to initiate NMD [17]. This suggests that normal stop codon are usually present at the last exon. This also explains why nonsense mutations occurring in the last exon and in transcripts lacking introns are able to escape NMD.

1.8 Introduction to microarray technology

Functional genomics attempts to answer questions about the functions of DNA at the level of genes, RNA transcripts, and protein products such as: monitoring the expression levels of a set of genes simultaneously under specific condition. This method of measurement of expression levels is called gene expression analysis. Microarray technology is one of the approaches to measure such expression levels and it involves analysis of large amounts of data through the use of various computational methods to derive meaningful results from such experiments [6]. The main principle behind microarray technology is, in situ hybridization of complementary sequences.

One of the major use of microarray in measuring gene expression levels is to compare expression of the same set of genes from cells maintained under one

condition (Condition A) to cells under normal condition (Condition B). RNA molecules extracted from the cells are reverse transcribed to cDNA where RNA from different conditions are labelled using two fluorochrome dyes Cy3 (green) and Cy5 (red) for two color microarray. For example cDNA from condition A may be labelled with red dye and cDNA from condition B with green dye. These cDNA are then allowed to hybridize on a single glass slide which then compete for binding to the probes fixed on the microarray surface. The hybridization step is followed by scanning of the chip twice, once for each wavelength and the final image is generated by overlaying previous two images. The final image summerizes an intensity reported for every spot for both wavelengths. The pixel intensity for each channel might be assigned a quality score, which allows to report poor quality spots: examples of spot quality assessment are the standard deviation of spot intensities, shape of array image, signal-to-noise ratio [19].

One of the quality plots used to assess the quality of raw microarray data is a MA plot. It is the plot of intensity ratio of raw microarray data where M, in Y axis, represents the log intensity ratio for red and green channel and A, in X-axis, represents the average log intensity. A good quality microarray data have elongated comet shaped MA plot with large number of genes at line M=0 representing equal intensities of the red and green channel. This suggests most of the genes in a microarray experiment are not differentially expressed [19].

The quality assessment is followed by normalization, which is an essential procedure to remove any systematic biases that do not represent any significant biological variations between samples [19]. There are several established normalization methods such as lowess regression (local weighted scatterplot smoothing) for intensity dependent normalization, which fits a weighted least square curve to the MA-plot and is subtracted from original log-ratio to obtain normalized log-ratio [19]. Similarly, another normalization approach is quantile normalization, which aims to make similar distribution of intensities across the set of arrays [19]. Overall, these normalization methods preserve the interesting variations while trying to remove unnecessary technical variations [19]. After normalization, the result of two color microarray experiments is reported as fold-change between channel intensities for each spot, which is the measure for mRNA abundance across two samples. The schematic experimental process of microarray is shown in Figure 1.4.

1.9 Transcriptional regulation and its complexity

The combination of transcription factors binding to promoter regions of genes regulate key functional processes of cells through regulation of gene expression. Thus, analysis of the promoter regions of genes expressed under specific conditions can be useful to describe disease-association of such genes based on the phenotypic properties [40].

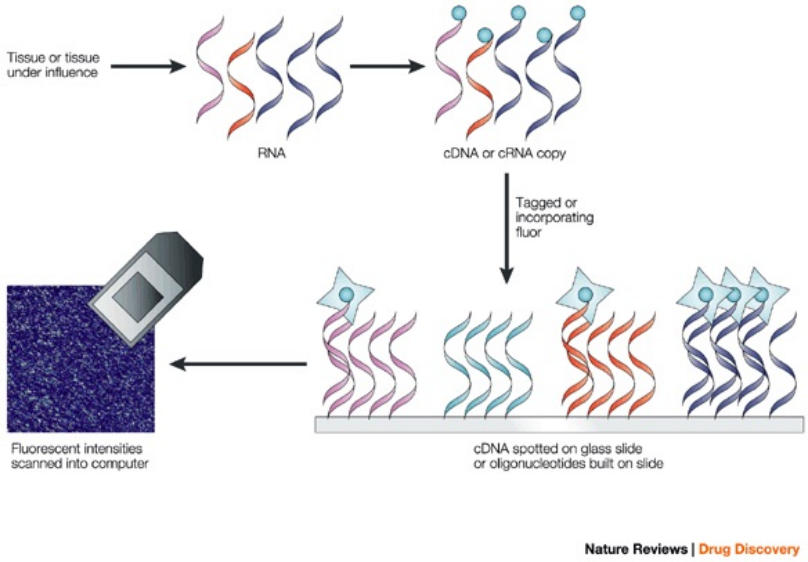


Figure 1.4: Representation of microarray experiment. Reproduced from [15] for the purpose of this discussion.

Transcription is a series of complex reactions, which is influenced by many regulatory signals. Furthermore, transcription is also context dependent as every regulatory signal is defined by unique affinity of DNA-binding molecules, transcription factors and histones. The interaction of these with DNA result in the formation of unique transcriptional machinery that leads to a transcriptional product [45].

Promoter region is upstream of the coding sequence where assembly of pre-initiation complex occurs, which helps to position RNA polymerase II complex at the transcription start site (TSS). The region where RNA polymerase II is recruited is called the core promoter and the rest of the region corresponds to transcription factor binding sites (TFBS). Absence of specific transcription factors at TFBS shuts down the activity of these promoter regions

in eukaryotes [45].

1.9.1 In Silico promoter analysis and prediction of regulatory motifs of functional relevance

The amino acid sequence of a protein is uniquely determined by the coding region of a gene. On the other hand, the regulatory sequences that are present at untranslated regions of the gene have indirect, non-linear and context dependent relationship with transcriptional regulation, for example in determining their transcription profile [45]. Regulatory regions play an important role in gene function and being able to predict them under specific condition will provide us with an additional insight on genotype-phenotype correlation. In most of these cases, co-regulation is the result of activation of common transcription factors interacting with cis-regulatory motifs present in their promoter region [45]. A regulatory pattern obtained from a group of transcription factor binding sites after alignment is usually represented as consensus sequence or a Position Weight Matrix (PWM).

1.9.2 Consensus sequence

It is the representation of most frequent residues of either nucleotide or amino acid calculated using distribution of those residues in the sequence alignment. Consensus sequence represents a conserved sequence motif in multiple sequence alignment and shows which residues are conserved and which residues are variable [52]. The motif is represented as letters of the IUPAC code, where the highest nucleotide composition of each column of the alignment is assigned for each position of the consensus sequence [45]. For example, in a consensus sequence notation for DNA sequence: TATAAT, each base occupies corresponding position even though the percentage of occurrence of 'T' and 'A' at all positions is not 100%. Thus, a consensus sequence does not take into account the relative frequency of occurrence of nucleotide. There exists an alternative method to represent a consensus sequence called *sequence logo*, which also considers the relative frequency of occurrence of nucleotides at that position. Sequence logo is the graphical representation of consensus sequence where the size of the symbol is directly proportional to the frequency of the nucleotide in the multiple sequence alignment. The more conserved residue has the larger symbol [52]. The consensus sequence of KLF4 motif is represented in Figure 1.5.

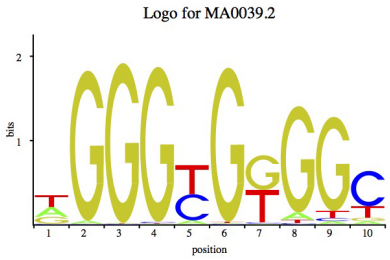


Figure 1.5: Sequence logo of KLF4 (MA0039.2) representing consensus sequence of KLF4 binding site in the promoter region.

1.9.3 Position Weight Matrix

It is a matrix constructed from a set of aligned sequences which are known to bind a common transcription factor, where each element is proportional to the observed count of nucleotide at the corresponding position [18]. The first step in construction of PWM from a set of aligned sequences is computing a matrix with a frequency of nucleotide at each position. The elements of resulting matrix can be interpreted as probability, by dividing each occurrence frequency by the number of sequences used to generate the matrix [27]. An illustrative example of calculation of PWM from set of aligned sequences is shown below. An example of aligned sequences from 5' $\rightarrow$ 3' is given below:

GAGGTAAAC
TAGGTCATT
AAGGTAAGT
CAGGTATAC
ACAGTCAGT
TCCGTAAGT
TGTGTGAGT

The frequency matrix for the above aligned a set of sequences is given by matrix M, where each row in matrix M represents frequency of the nucleotides A, C, G, T respectively in the corresponding position of the alignment represented by columns in the matrix M. For instance, M[1,1]=2 represents the frequency of A at position 1 of the sequence.

$$M = \begin{bmatrix} 2 & 4 & 1 & 0 & 0 & 4 & 6 & 2 & 0 \\ 1 & 2 & 1 & 0 & 0 & 2 & 0 & 0 & 2 \\ 1 & 1 & 4 & 7 & 0 & 1 & 0 & 4 & 0 \\ 3 & 0 & 1 & 0 & 7 & 0 & 1 & 1 & 5 \end{bmatrix}$$

Similarly, above matrix M can be modified to represent the probability of nucleotide occurrence by dividing each element of M by the number of sequences (N=7).

$$M = \begin{bmatrix} 0.29 & 0.57 & 0.14 & 0 & 0 & 0.57 & 0.86 & 0.29 & 0 \\ 0.14 & 0.29 & 0.14 & 0 & 0 & 0.29 & 0 & 0 & 0.29 \\ 0.14 & 0.14 & 0.57 & 1 & 0 & 0.14 & 0 & 0.57 & 0 \\ 0.43 & 0 & 0.14 & 0 & 1 & 0 & 0.14 & 0.14 & 0.71 \end{bmatrix}$$

PWM are generally transformed using a background model and represented as log likelihoods [27]. A simplest background model is to assume equally likely occurrence of nucleotides in the data, which would be 0.25 (4 different DNA bases) for nucleotides and 0.05 (20 different amino acids) for amino acids. The log likelihoods for PWM is calculated as:

$$M = \ln\left(\frac{M_{i,j}}{B}\right) \tag{1.1}$$

Transformation of above matrix M using Equation 1.1 gives:

$$M = \begin{bmatrix} 0.13 & 0.83 & -0.56 & \infty & \infty & 0.83 & 1.23 & 0.13 & -\infty \\ -0.56 & 0.13 & -0.56 & \infty & \infty & 0.13 & -\infty & -\infty & 0.13 \\ -0.56 & -0.56 & 0.83 & 1.39 & -\infty & -0.56 & -\infty & 0.83 & -\infty \\ 0.54 & \infty & -0.56 & \infty & 1.39 & -\infty & -0.56 & -0.56 & 1.05 \end{bmatrix}$$

In order to avoid the value of '-∞' in PWM generated from a small number of sequences, pseudocounts are often applied while calculating PPMs [44]. Furthermore, background model can also be adjusted according to the genomic region. For example, a frequency of bases G and C can be empirically increased for GC rich region. Using these normalized frequencies of the four nucleotides, PMW model reflects binding preferences at each position. Thus, a score can be calculated using the values observed at each corresponding nucleotide position which is proportional to the binding energies [45]. Several databases are available which record the PMWs of specific transcription factors such as JASPER, TRANSFAC. On the other hand, pattern discovery detects common motifs in a group of unaligned sequences. There exist many methods for prediction of relevant TFBSs in a set of sequences and are based on word counting, pattern matching (PWM) or pattern discovery algorithm. Although prediction of single TFBS might have a higher false positive rate, predicted overrepresentation of TFBS are likely to have functional relevance as this method is highly dependent on the background model [45].

1.10 HSPB1 mutation explored in this study

In this study, I investigate a novel heterozygous HSPB1 mutation in a patient with dominantly inherited motor predominant axonal Charcot-Marie-Tooth disease: a unique frame shift mutation predicting ablation of the entire C-terminal domain of the protein p.(M169Cfs2X) and a missense mutation HSPB1^{R127L}. Similar truncated variants were also described earlier, but the stability of putative truncation was not investigated [35, 50]. Thus, the mechanism of these truncations remained obscure. Furthermore, missense mutation HSPB1^{R127L} is studied together with the truncated variant, but dropped later in the study as fibroblasts of this missense variants behaved normally and produced no clear phenotypes. In vitro study of mutant protein was carried out using patient derived primary fibroblast. This study demonstrates the stability of truncated HSPB1 and its ability to bind its wild type counterpart, suggesting a dominant negative effect, which impairs the ability of patient cells to cope with the stress.

1.11 Aims of this study

The identification of a novel heterozygous mutation (p.M169Cfs2X) predicting ablation of entire C-terminus of the protein HSPB1 provided us with an opportunity to investigate the cellular, transcriptional and biochemical consequences of mutation in patient derived primary fibroblasts as the disease model. Through this project, I attempted to find out answers for at least some of the questions listed below:

1. Is the truncated protein stable in patient fibroblasts and is the dimerization affected?
2. Do the truncated proteins translocate to the nucleus upon heat stress?
3. Does the truncated protein affect the transcriptional heat stress response?
4. Are the patient cells sensitive to heat stress?
5. What is the role of HSPB1 in transcriptional heat stress response?
6. Does the transcriptional profile of the HSPB1 patient cells differ from that of control cells?
7. Are there common transcription factors co-regulating genes of specific pathway in up-regulated and down-regulated genes?

Chapter 2

Materials and Methods

After identifying the novel truncated HSPB1 (HSPB1$^{\triangle \text{C-term}}$) mutant, we were interested in studying its functional effect and pathways that were altered in the patient. The first step was to elucidate the stability of the truncated HSPB1 followed by other experimental and computational methods. This section discusses the details of the methods used in the project. The general workflow of the project is outlined in the Figure 2.1

2.1 Cell Culture

Control and patient (HSPB1$^{\triangle \text{C-term}}$) fibroblasts were cultured at 37°C in DMEM supplemented with 10% FBS, L-glutamine, Pen-Strep, and Uridine in the presence of 5% CO_2 in all experiments unless specified.

2.2 Western Blotting

For western blotting, lysis buffer RIPA (1x PBS, 1% Nonidet-P-40, 0.5% sodium deoxycholate, 0.1% SDS) was used to prepare whole-cell lysates. Cells were suspended in 50 Mm Tris-HCl (pH 8.0), 10% glycerol, 1% Nonidet P-40, 150 mM NaCl, 5mM NaF, 5uM $ZnCl_2$, 1 mM Na_3VO_4, 10mM EGTA and complete protease inhibitor. Then, the cells were lysed on ice for 10 minutes and centrifuged and pellet discarded. The supernatant was boiled in non-reducing loading buffer (250 mM Tris-HCl, pH 6.8, 10% SDS, 30% glycerol, 0.02% bromophenol blue) for 5 minutes. For nuclear enrichment, cells were first homogenized by passing through 22G syringe needle 8-12 times in a buffer (0.3M sucrose, 1mM EGTA, 5mM MOPS, 5mM KH_2PO_4 and complete protease inhibitor) at pH 7.4 followed by centrifugation at 6500 rpm

for 15 minutes at 4°C. The supernatant and pellet were collected as cytosolic fraction and nuclear fraction respectively. Furthermore, nuclear fractionations were done for both non-treated cells and cells subjected to 45°C heat shock for 1 hour. Similarly, for extracellular HSPB1, patient and control fibroblasts were grown in 6 well plates for 24 hours and 50 μl of the medium from the plates were transferred to 1.5 ml centrifuge tubes and centrifused at 1500 rpm for 15 minutes. Supernatants from each tube were transferred to new ones and stored at -80°C until use. Proteins were separated by SDS-PAGE following the standard protocol.

Before transfer of proteins to a nylon membrane, the membrane was pretreated with methanol for 10 seconds and washed with milliQ water for 2 minutes and kept in semi-blot solution (2.9 g glycine, 5.8 g Tris-base, 3.75 ml 20% SDS in 800 ml water and 200 ml methanol). For transfer of proteins from gel to membrane; three Whatman paper, SDS-PAGE gel, nylon membrane followed by another 3 Whatman paper were stacked in transfer apparatus and ran for 1 hour and 10 minutes at 50 mA (milliampere). After completion of the transfer, the membrane was blocked with 5% milk for 1 hour and incubated with primary antibody overnight in a shaker at 4°C. After the incubation with primary antibody, the membrane was washed with Tris-buffered saline with Tween 20 (TBST) solution 3 times in a shaker for 15 minutes, followed by incubation with secondary antibody for 1 hour at room temperature and washing with TBST as before.

For visualization, the membrane was incubated in dark with 1.5 ml of substrate (peroxide solution) and 1.5 ml of enhancer (luminol) for 5 minutes and imaged using BioRad Chemidoc$^{\text{TM}}$.

Antibodies used were: anti-HSPB1 (18284-1-AP; Proteintech, Chicago, IL, USA), anti-GAPDH, anti-histone H3 and anti-tubulin (2118, 4499 and 2146 respectively, Cell Signaling Technology, Danvers, MA, USA). The general experimental set up for protein transfer in western blot is shown in Figure 2.2.

2.3 Immunocytochemistry

Control and patient (HSPB1$^{\triangle\text{C-term}}$) fibroblasts were cultured according to a standard method. Media was removed and the cells were rinsed 3 times with phosphate-buffered saline (PBS). Then were fixed on glass slides with 4% paraformaldehyde for 10 minutes and again rinsed with PBS 3 times.

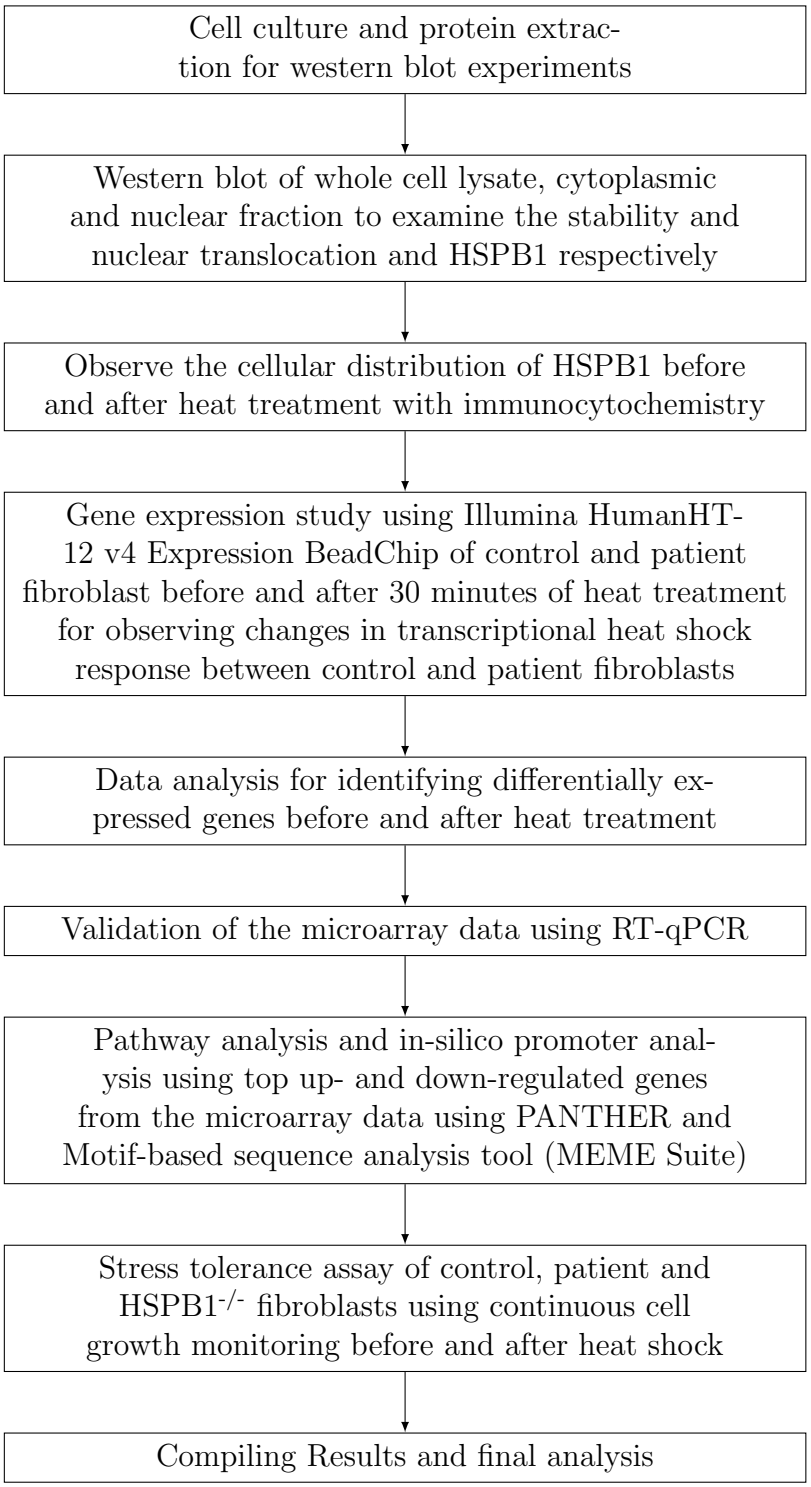


Figure 2.1: Work flow of the project

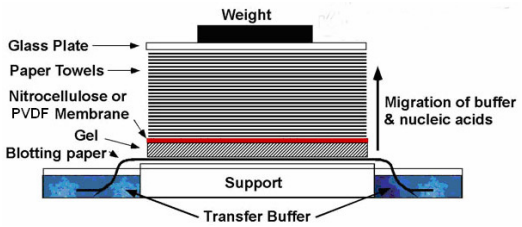


Figure 2.2: A set up for protein transfer in western blot. Source [46]

Cellular permeabilization was done by incubating the cells for 15 minutes with 0.2% Triton X-100 and were washed three times with phosphate-buffered saline with Tween 20 (PBST) (0.1% Tween 20). To prevent unspecific binding of antibody, the slides were blocked with 3% BSA/PBST for 2 hours at room temperature. Primary antibody dilutions were prepared in 3% BSA/PBST at a ratio of 1:1000. The cells were incubated with primary antibody at room temperature for 12 hours in a shaker and were covered with foil to prevent drying. The slides were then washed three times with PBST for 10-15 minutes each. Furthermore, secondary antibody dilutions were also prepared in 3% BSA/PBST in the ratio 1:2000 and incubated at room temperature for 1 hour and washed three times with PBST for 10-15 minutes each. After final wash, cells were mounted with 30μl VECTASHIELD® and coverslips were dropped on top of mounting media and left for 1 hour in dark to dry. Same antibodies were used as in western blotting and DAPI was used for nuclear staining. The cells were visualized using Axioplan 2 fluorescent imaging unit.

2.4 Gene Expression Microarray

Control and patient (HSPB1$^{\triangle \text{C-term}}$) fibroblasts were subjected to heat shock for 30 minutes at 45°C and total RNA was extracted from both treated and untreated cells using Nucleospin® total RNA extraction kit according to the protocol specified by the manufacturer using three independent replicates. The gene expression profiling was performed using Illumina HumanHT-12 v4 BeadChip. The files obtained from GenomeStudio were read by R using *lumiR()* function from *lumi* package [21] and annotated using latest annotation file from Illumina. The data were corrected for background adjustment and normalized using quantile normalization method implemented in function *lumiExpresso()* of *lumi* package. GenomeStudio outputs intensity values for each probe together with the detection p-value. Probes with detection p-value ≤ 0.01 were selected for further analysis. The details of the design of

the microarray experiment is shown in the Table 2.1. We were interested in realizing the differential expression between patient and control at untreated condition, P0-C0 and at treated condition, P30-C30. Similarly, response of control to treatment (C30-C0) and response of the patient to treatment (P30-P0) were also evaluated. Test for differential expression was performed in R using *lmFit()* function implemented in *limma* package [49] followed by *topTable()* function, with sort by 'none' option to return the list of all genes together with their associated log ratio and adjusted p-value. By default the method of Benjamini-Hochberg was used to correct for multiple testing, by calling the *p.adjust()* function.

Array Number	Sample Name	Sample Type	Explanation
9989522057_A	P_0_1	P0	Patient, no heat shock
9989522057_B	P_0_2		
9989522057_C	P_0_3		
9989522057_D	C_0_1	C0	Control, no heat shock
9989522057_E	C_0_2		
9989522057_F	C_0_3		
9989522057_G	P_30_1	P30	Patient, with 30 min heat shock
9989522057_H	P_30_2		
9989522057_I	P_30_3		
9989522057_J	C_30_1	C30	Control, with 30 min heat shock
9989522057_K	C_30_2		
9989522057_L	C_30_3		

Table 2.1: Microarray experiment design.

2.5 Gene set enrichment analysis for differentially expressed genes

This test was performed on the list of top differentially expressed genes in order to identify if a particular class of genes (e.g., Gene Ontology biological processes or the PANTHER pathways) is overrepresented or underrepresented. PANTHER program [37] was used to perform this analysis and was carried out on genes with either two fold up- or down-regulation of gene expression and p-value ≤ 0.01, which is equivalent to log-ratios of +1 or -1 respectively. The reference list used for the analysis contained all the genes from Illumina HumanHT-12 v4 BeadChip. The PANTHER program then applied the binomial test for each PANTHER class to determine whether

there is an overrepresentation or underrepresentation of genes in the test set relative to the reference set. The 'expected value' in the analysis was the number of genes that would be present in the test list based on the reference list. For instance, among approximately 47,000 genes in Illumina HumanHT-12 v4 BeadChip, 5,000 maps to GO term cell cycle. Therefore, 0.1% (5,000 divided by 47,000) genes in reference list were involved in cell cycle. If the test list uploaded to the PANTHER contains 10,000 genes, 10 genes (10,000 multiplied by 0.1%) would be expected to be involved in cell cycle. Now, for any biological process, if more genes were observed in the test list than expected, there was an 'overrepresentation' (+) of genes involved in the process. If fewer genes were observed than expected, there was an 'underrepresentation' (-). A p-value was calculated to determine if the overrepresentation (+) or underrepresentation (-) was significant or not. The statistical method used to perform the overrepresentation test was a binomial test.

2.6 In-Silico promoter analysis for identification of over-represented motifs in upregulated and downregulated genes

The aim of this analysis was to identify the putative upstream regulatory elements that might have caused the changes in the gene expression levels observed between the patient (HSPB1$^{\triangle \text{C-term}}$) and the control fibroblasts.

The first step in this analysis was to filter the up-regulated and down-regulated genes independently based on the adjusted p-value and fold change cut-off. The data was generated to capture early responding genes to heat shock. Thus, it is expected that those genes would vary widely among the other genes. However, we also wanted to capture genes with intermediate changes for the promoter analysis, as it is known that not all mRNAs are translated into proteins. Similarly, some mRNAs are more stable compared to others, thus requiring less transcription [11]. These motivated us to set a less stringent criterion to filter genes for promoter analysis. I chose adjusted p-value cut-off ≤ 0.01 and log ratio ≥ 1.75 for up-regulated genes and log ratio $\leq$ -1.75 for down-regulated genes. The promoter sequences of all genes in the list were extracted using EPD database [20], as it contains experimentally determined transcription start site (TSS) and is an annotated non-redundant database of eukaryotic POL II promoters. For all the genes, sequence between -499 to 100 bp relative to TSS was extracted, which corresponds to transcription factor binding sites (TFBS) at short distance from

the TSS. For genes with splicing variants, promoter sequences from all variants were extracted, as we had no information regarding the variant level expression of genes from the microarray data. Motif enrichment analysis (MEA) was implemented using the AME tool [36] and CentriMo algorithm [7] of the MEME suite against known TF matrices in JASPER CORE (2014) vertebrates database.

The MEA method is based on the expectation maximization (EM) algorithm. The algorithm uses motif model represented as a position weight matrix (PWM) and both the motif model and its locations in the sequences are optimized iteratively. The choice of background affects the accuracy of the AME algorithm [36]. I performed the analysis with shuffled input sequences as control sequences for the analysis. Furthermore, I performed the same analysis with CentriMo, which uses PWM to scan the set of equal-sized peak regions (promoter sequences). It declares at most one binding site in each equal-sized peak region and discards regions with no declared site. Thus, each remaining region contains one site and the null model assumes the site to be uniformly distributed within the region. This assumption implies that a binomial model can be applied to the number of sites in the central region. This will account for the background choices as flanking sequence can be used as background model.

2.7 Total RNA extraction and RT-qPCR for validation of micorarray data

Total RNA from 3 controls, patient (HSPB1$^{\triangle \text{C-term}}$) and HSPB1$^{-/-}$ fibroblasts was extracted using the same kit used for total RNA extraction for microarray experiment according to the protocol specified by the manufacturer. The knockout cell line of HSPB1 was generated using CRISPR-Cas genome editing vector for HSPB1 provided by Horizon's free CRISPR Guide RNA program. Concentration of total RNA was measured using NanoDrop spectrophotometer. Among the top up-regulated and down-regulated genes, 11 genes were randomly selected for validation through RT-qPCR. Primers for all the genes were designed using the Primer-Blast tool from NCBI.

2.7.1 cDNA synthesis

530 ng of total RNA was used as a template for the reverse transcription in 20μl volume. The first step was to incubate total RNA with random

hexamer primer for 5 minutes at 65°C. The combined volume of water and RNA (x and y in Table 2.2) had 530 ng of total RNA. This incubation was followed by addition of 10μl of reverse transcription buffer and 2μl of reverse transcription mix (Thermo Scientific DyNAmo cDNA Synthesis kit, F-470L) and running a PCR according to the conditions mentioned in Table 2.3.

	Volume (μl)
Water	x
RNA	y
Random Primer	1
Total	8

Table 2.2: Random Priming

Temperature	Time (Minutes)
25°C	10
37°C	60
85°C	5

Table 2.3: First strand synthesis

2.7.2 RT-qPCR

The template from above cDNA synthesis step was diluted to get 25 ng of cDNA in 5 μl of water. The composition of the reaction mixture and reaction conditions are shown in Table 2.4 and Table 2.5 respectively.

	Volume (μl/well)
cDNA	5
2× Syber	10
F-Primer	1
R-Primer	1
Water	3

Table 2.4: RT-qPCR Mix

Temperature	Time	Cycles
95°C	7 min	
95°C	10s	↑
60°C	30s	

Table 2.5: Reaction condition. Blue color and arrow indicate the cycle is repeated 40 times.

2.8 RT-qPCR of heat shock genes in control, HSPB1$^{\triangle \text{C-term}}$ and HSPB1$^{-/-}$ fibroblast before and after heat shock

To assess the affect of HSPB1 in the transcriptional heat response in patient fibroblasts, it was essential to have a cell model with total knockout of HSPB1. Total RNA from HSPB1 knockout fibroblasts along with three control fibroblasts and patient (HSPB1$^{\triangle \text{C-term}}$) fibroblasts was extracted at normal condition and after 1 hour of heat treatment at 45°C using Nucleospin®

total RNA extraction kit according to the protocol specified by the manufacturer. As discussed in the Section 1.6, HSPB1 does not participate in the active refolding of proteins, but can act as a substrate holder for ATP dependent HSPs such as HSP70. DNAJB1 is known to interact with HSP70 and stimulate its ATPase activity and HSPA6 is a homolog of HSP70. Thus, HSPB1 might influence their activity. These genes (HSPA6 and DNAJB1) were selected to test the role of HSPB1 in transcriptional heat shock response, as these genes were also seen to be up-regulated in both control and patient during heat shock in the microarray experiment. Primers for both genes were designed using Primer-Blast from NCBI. For the RT-qPCR experiment, GAPDH was used as reference gene as it had a stable expression in ($HSPB1^{\triangle C\text{-}term}$), control and $HSPB1^{-/-}$ fibroblasts. 1 mg of total RNA was used as a template for cDNA synthesis and RT-qPCR. Both were performed using the protocol specified in Subsection 2.7.1 and 2.7.2.

2.9 Stress Tolerance Assay

Control, patient ($HSPB1^{\triangle C\text{-}term}$) and $HSPB1^{-/-}$ fibroblasts were cultured at 37°C and 5% CO_2 in a 12 well plate with triplicates at a cell density of 6.6 x 10^3cells/well. The medium used was DMEM supplemented with 10% FBS, L-glutamine, Pen-Strep, and Uridine. The cells were allowed to grow for 24 hours and moved to Cell-IQ system (CM Technologies, Tampere, Finland) and followed for 24 hours with 49 regions imaged every 3 hours. Medium was changed and cells were subjected to heat shock for 2 hours at 45°C and returned to normal condition and followed for 24 hours with the same 49 regions imaged every 3 hours. The images were analyzed using the Cell-IQ image analysis software provided by CM Technologies. The software was trained manually to detect fibroblast cells in the image. Total cells, before and after heat shock, were counted at each time point to observe the growth curve of cells. Statistical significance in growth rate between cell types was tested using two-way ANOVA.

Chapter 3

Results

This chapter mainly focuses on the description of the results obtained from the experimental methods described in Chapter 2.

3.1 Truncated HSPB1 is stable and is able to dimerize with its wild type counterpart

Errors in gene expression that results in mRNA transcripts with premature stop codons are eliminated by a NMD pathway discussed in the Section 1.7. To assess if the truncated *HSPB1* is stable or not, I ran a SDS-PAGE on lysates of primary patient fibroblasts. Although patient fibroblast with missense mutation HSPB1$^{\mathrm{R127L}}$ had similar amounts of full length *HSPB1*, HSPB1$^{\triangle\mathrm{C\text{-}term}}$ fibroblast showed two bands while reacting with anti-HSPB1. The two bands in HSPB1$^{\triangle\mathrm{C\text{-}term}}$ fibroblast correspond to full length ($\sim$27 kDa) and truncated *HSPB1* ($\sim$23 kDa) respectively as shown in Figure 3.1.

It was known to us that *HSPB1* dimers are resistant to denaturation. However, they dissociate when treated by reducing agents like SDS [4]. Thus, to realize if truncated *HSPB1* could form homodimers and heterodimer with itself and wild type, I performed western blots under non-reducing conditions that would allow us to visualize oligomers of *HSPB1*.

According to protein bands observed in the Figure 3.2, the control fibroblasts showed bands that correspond to the wild type $\sim$27 kDa monomer and the wild type $\sim$50 kDa dimer. However, in the same blot the HSPB1$^{\triangle\mathrm{C\text{-}term}}$ fibroblasts showed two additional faint bands between the normal sized monomer and dimer. These bands can be explained as heterodimers formed by a wild type and a mutant protein, and homodimers formed by two mutant pro-

teins respectively. Furthermore, from Figure 3.2 we can also distinguish that the total abundance of wild type *HSPB1* homodimer was relatively less in HSPB1$^{\triangle\text{C-term}}$ fibroblasts when compared to control fibroblasts. This result suggests HSPB1$^{\triangle\text{C-term}}$ protein binding to wild type *HSPB1* protein exerted a possible dominant negative effect resulting in a decreased abundance of wild type homodimer [58].

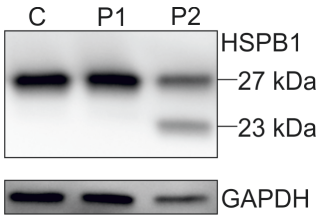


Figure 3.1: Truncated *HSPB1* was stable under reducing conditions as shown by western blot followed by immunodetection of RIPA lysates of primary fibroblasts with anti-HSPB1 antibody. The figure is a representative blot from control (C), HSPB1$^{\text{R127L}}$ (P1) and HSPB1$^{\triangle\text{C-term}}$ (P2) fibroblasts respectively. HSPB1$^{\triangle\text{C-term}}$ (P2) fibroblasts shows two bands ∼27 kDa corresponding to wild type *HSPB1* and ∼23 kDa corresponding to truncated *HSPB1*. *GAPDH* was used as a loading control.

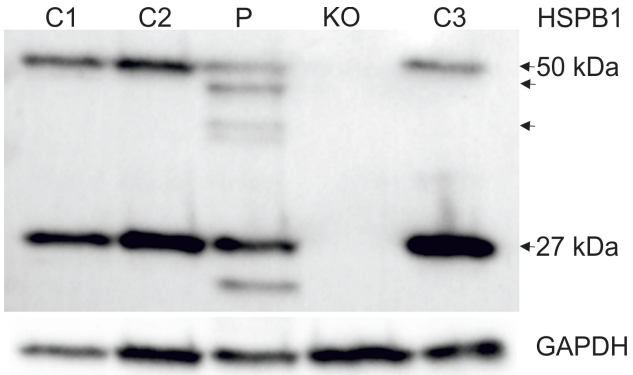


Figure 3.2: Western blots under non-reducing conditions. For control fibroblasts (C1-3) a band at ∼27 kDa represents the wild type monomer and a band at ∼50 kDa represents the wild type dimer. In HSPB1$^{\triangle\text{C-term}}$ fibroblasts (P), in addition to above two bands, two additional fainter bands (arrowheads) can be observed, which represents heterodimers formed by wild type and truncated HSPB1 and homodimers formed by two truncated *HSPB1*. The *HSPB1* knockout fibroblast (KO) does not show any bands corresponding to HSPB1. *GAPDH* was used as loading control.

3.2 Truncated HSPB1 (HSPB1$^{\triangle \text{C-term}}$) translocates normally to nucleus

In the cells stressed with mild heat (43°C), *HSPB1*, previously known as *HSP25* or *HSP27*, changes its cellular location, forming characteristic granules in the nucleus [14]. To observe if there was any nuclear translocation activity of truncated *HSPB1*, I used western blot of nuclear enriched fraction after heat shock and immunocytochemistry methods as described in Chapter 2. From the Figure 3.3 (A) it was clear that *HSPB1* translocates to the

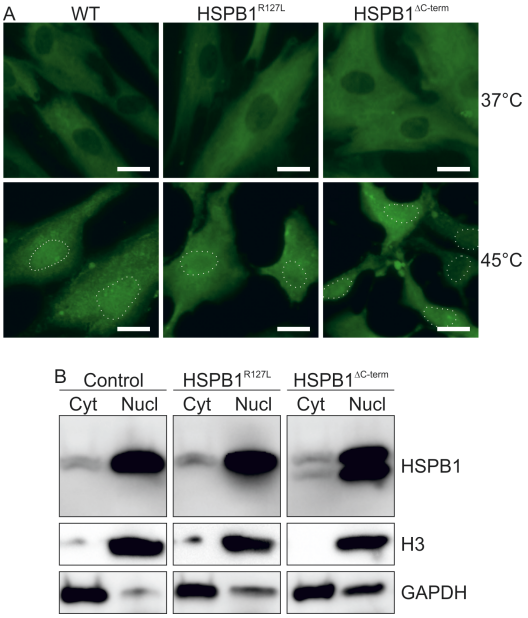


Figure 3.3: *HSPB1* translocated to the nucleus under heat stress. Figure A is a representative immunocytochemistry of primary fibroblasts with anti-HSPB1 antibody (green) and Figure B is a western blot performed after nuclear enrichment. It can be observed that, fraction of cytoplasmic *HSPB1* translocated to the nucleus at 45°C in control (WT), HSPB1$^{\triangle \text{C-term}}$ and HSPB1$^{\text{R127L}}$ fibroblasts. Nuclei are represented as dotted lines and Bars = 20 μm. (B) After Nuclear enrichment and Western blotting, there was increased abundance of *HSPB1* in the nucleus (Nucl) compared to the cytoplasm (Cyt) at 45°C. Histone H3 protein was used as a nuclear loading control, and *GAPDH* as a cytoplasmic loading control. It can be realized from the image that the HSPB1$^{\triangle \text{C-term}}$ protein translocates to the nucleus and is detectable after heat stress.

nucleus upon heat stress as the nuclear region (shown in dotted circles) were heavily stained in green after heat stress. Similarly, the presence of truncated *HSPB1* in the nuclear fraction from the western blot confirmed that both wild type and truncated *HSPB1* translocated to the nucleus. From the

Figure 3.3 (B), it is evident that truncated *HSPB1* is enriched in the nuclear fraction of patient fibroblasts (HSPB1$^{\triangle \text{C-term}}$). These results support that there is no defect in nuclear translocation of *HSPB1* in patient fibroblasts. Furthermore, cytoplasmic and nuclear *HSPB1* concentration, *HSPB1* nuclear translocation upon heat stress of HSPB1$^{\text{R127L}}$ fibroblasts were similar to that of control. Thus, these fibroblasts were not included in further analysis.

3.3 Gene expression analysis of control and patient (HSPB1$^{\triangle \text{C-term}}$) fibroblasts

The results from nuclear translocation of truncated *HSPB1* raised an important question; does it influence global transcriptional heat response in patient fibroblasts? To answer this, we decided to perform a gene expression analysis between control and patient fibroblasts. The experimental design for the microarray was discussed in Chapter 2. The results of the differential expression analysis of patient vs. control without heat treatment (P0-C0) and patient vs. control with heat treatment (P30-C30) at different p-value and log ratio cut-offs are shown in the Table 3.1. We observed 49 and 47 differentially expressed genes for P0-C0 and P30-C30 at log ratio cut-off between -2 and 2 respectively. There were 44 common genes differentially expressed between two conditions. Similarly, there were 73 and 71 differentially expressed genes at log ratio cut-off between -1.75 and 1.75 with 68 common genes differentially expressed between two conditions. This result suggests similar global transcriptional profile of patient vs. control under normal (P0-C0) and 30 minute heat treatment (P30-C30) condition, as there were a similar number of genes differentially expressed and most of the differentially expressed genes were common to both conditions.

Cut-off	P0-C0 (A)	P30-C30 (B)	A ∩ B
p-value ≤ 0.01	4050	2472	2241
p-value ≤ 0.01, $-2 \geq \text{logFC} \geq 2$	49	47	44
p-value ≤ 0.01, $-1.75 \geq \text{logFC} \geq 1.75$	73	71	68

Table 3.1: Numbers of significantly differentially expressed genes with different cut-off values between the two comparisons. A represents the condition P0-C0, B represents condition P30-C30 and A ∩ B represents number of common differentially expressed genes between two conditions.

3.4 Pathway enrichment

Pathway analysis was performed on genes with either two fold up- or down-regulation in gene expression and p-value ≤ 0.01, which is equivalent to log-ratios of +1 or -1 respectively. This analysis was performed with an assumption that genes with small changes in expression levels can have affect at the pathway level through synergy. As discussed in the Section 2.5, pathway analysis using PANTHER revealed enrichment of negative regulation of apoptotic processes, macrophage activation, cellular defence response and immune system process. Enrichment of these pathways suggests early response of cells to stress and effect on growth of cells. As expected, our analysis identified the same enriched pathways both in the treated and in the untreated condition because the top genes under both conditions had high overlap. The Table 3.2 shows pathways enriched only under untreated condition (P0-C0) along with their fold enrichment and p-value.

An interesting finding was enrichment of extracellular matrix proteins as shown by enrichment of cell-cell adhesion. Few previous studies have provided data with the extracellular release of *HSPB1*, its possible involvement in immunogenic function and its ability to chaperone antigenic peptides [33, 47]. Furthermore, increased serum *HSPB1* was observed in cancer patients [33]. However, the role of soluble *HSPB1* was not discussed. Similarly, it was also reported that *HSPB1* secretion was increased in macrophages by estrogen and *HSPB1* was found within secretary lysosomal-like vesicles [47]. In our case, the presence of two kinds of *HSPB1* in patient cells raise further questions such as: Can both truncated and wild type *HSPB1* be secreted? If yes, how do they affect the normal function? I was not able to investigate these questions since I could not detect the secreted *HSPB1* from conditioned medium by the western blot following the described protocol [33].

3.5 Role of HSPB1 in transcriptional heat response

To evaluate the response of control (C0-C30) and response of patient (P0-P30) fibroblasts to heat treatment I analyzed the differences in the expression levels of genes after heat shock for both the patient and the control samples. The result of the analysis is shown in Table 3.3. This analysis resulted in 5 differentially expressed genes at p-value ≤ 0.01 and $-1.75 \geq \text{logFC} \geq 1.75$ under both condition suggesting patient and control fibroblasts have similar

transcriptional heat shock response. The 5 common differentially expressed genes were encoding proteins for heat shock and stress response (*DNAJB1, EGR1, FOS, HSPA6* and *HSPA7*).

Pathway	Genes in Background	Genes in Test	Expected	Fold Enrichment	p-value
macrophage activation	264	17	4.28	3.97	4.36E-04
negative regulation of apoptotic process	182	11	2.95	3.73	4.79E-02
cell-cell adhesion	494	27	8.01	3.37	1.15E-05
cellular defense response	372	19	6.03	3.15	2.94E-03
cell adhesion	858	39	13.91	2.8	1.68E-06
response to stimulus	1590	56	25.77	2.17	5.74E-06
immune system process	1647	56	26.69	2.1	1.87E-05

Table 3.2: Enriched pathways obtained after pathway analysis of gene-set from top up- and down-regulated genes of Patient Vs Control without heat treatment (P0-C0).

As discussed in Section 2.8, *HSPB1* does not actively refold misfolded proteins, but can act as a substrate holder for ATP dependent HSPs such as *HSP70*. Since *DNAJB1* and *HSPA6* are ATP dependent HSPs, absence of *HSPB1* might influence their activity. To further evaluate the role of *HSPB1* in transcriptional response, the cells were exposed to heat stress for a longer period of time (1 hour) and expression levels of the two heat shock proteins: *DNAJB1* and *HSPA6* were quantified using RT-qPCR. From the Figures 3.4 and 3.5, it can be observed that although relative expression levels of both genes in HSPB1$^{-/-}$ cells are lower compared to control, it is not significant. This suggests *HSPB1* might not have participated in early heat shock response of patient fibroblasts.

3.6 Patient and HSPB1$^{-/-}$ fibroblast are sensitive towards heat stress

As described in Section 2.9, it was possible to assess the effect of heat treatment on the growth rate of control, patient (HSPB1$^{\triangle \text{C-term}}$) and HSPB1$^{-/-}$ fibroblasts using continuous video monitoring through Cell-IQ system. The images captured immediately after heat shock were out of focus hence they were not included in the analysis. The number of live cells before and after heat shock of control, patient (HSPB1$^{\triangle \text{C-term}}$) and HSPB1$^{-/-}$ fibroblasts revealed significant difference in growth rate between control and patient (HSPB1$^{\triangle \text{C-term}}$) fibroblasts and control and HSPB1$^{-/-}$ fibroblasts. However,

the difference in growth of patient (HSPB1$^{\triangle \text{C-term}}$) and HSPB1$^{-/-}$ fibroblasts was significant only at 24 hours time point. The growth curves of all three cell lines before and after heat shock are shown in Figure 3.6.

Cut-off	C0-C30 (A)	P0-P30 (B)	A ∩ B
p-value $\leq$ 0.01	50	5	5
p-value $\leq$ 0.01, -1.75 $\leq$ logFC $\leq$ 1.75	5	5	5

Table 3.3: Numbers of significantly differentially expressed genes with different cut-off values between the two comparisons. A represents, response of control (C0-C30), B represents a response of the patient (P0-P30) and A ∩ B represents the number of common genes between two conditions.

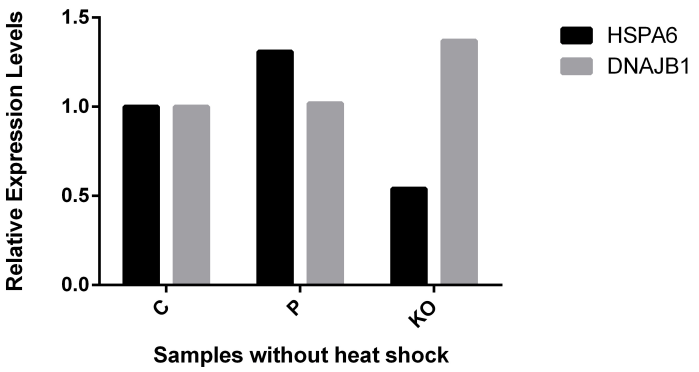


Figure 3.4: Relative mRNA level of HSPA6 and DNAJB1 before heat treatment for control (C), patient (P) and HSPB1$^{-/-}$ (KO). Data normalized to the expression level of control (C) before heat treatment.

3.7 Analysis of putative transcription factor binding sites

Since the patient cells and HSPB1$^{-/-}$ cells were sensitive towards heat stress, but expression levels of heat shock genes (*HSPA6* and *DNAJB1*) were similar to control cells after heat shock, the question was how the growth rate of patient cells and HSPB1$^{-/-}$ cells were affected? To answer this question, one idea was to analyse the promoter sequences of up-regulated and down-regulated differentially expressed genes. This would allow us to identify putative transcription factor binding sites in promoter region of these genes that might

co-regulate genes responsible for cell growth.

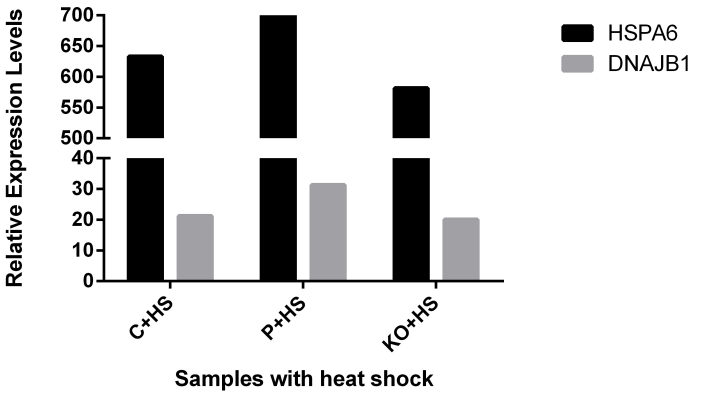
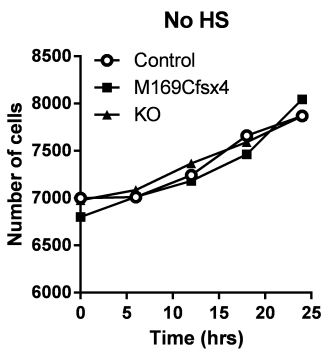
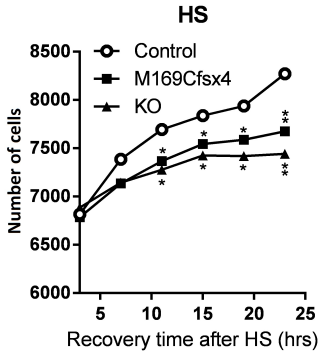


Figure 3.5: Relative mRNA level of HSPA6 and DNAJB1 after 1 hour of heat treatment for control (C), patient (P) and HSPB1$^{-/-}$ (KO). Data normalized to the expression level of control (C) before heat treatment.

Transcription factor motif enrichment analysis through AME and Centrimo was performed on the experimentally validated promoter sequences of top up- and down-regulated differentially expressed genes from both untreated (P0-C0) and 30 minutes heat treated (P30-C30) conditions independently as described in Section 2.6. The analysis identified enrichment of possible binding sites of four transcription factors *SP1*, *SP2*, *Klf4* and *Klf5* within 180 bp distance in the promoter site of down-regulated genes in untreated samples (P0-C0). However, in the promoter region of upregulated genes both AME and Centrimo were not able to find any significantly enriched motifs other than a TATA box (TBP), which is common in all promoter sequences. There were 73 differentially expressed genes under untreated condition (P0-C0) and 71 differentially expressed genes under 30 minute heat treated condition (P30-C30) with 68 genes common between them. As expected, we observed enrichment of same transcription factor motifs in down-regulated genes in treated samples (P30-C30) because of the large overlap in differentially expressed genes. The results of the motif enrichment analysis for untreated (P0-C0) condition is summarized in Table 3.4. The predicted occupancy site for these four transcription factors within the 600 bp region of the promoter sequences of down-regulated genes is shown in the Figure A.1.



(a) Growth curve of control, M169CfsX4 (HSPB1$^{\triangle \text{C-term}}$) and KO (HSPB1$^{-/-}$) fibroblasts under normal condition.



(b) Growth curve of control, M169CfsX4 (HSPB1$^{\triangle \text{C-term}}$) and KO (HSPB1$^{-/-}$) fibroblasts after 1 hour of heat shock at 45°C followed by 24 hours of recovery under normal condition.

Figure 3.6: Response of control, patient and HSPB1$^{-/-}$ fibroblast (a) before and (b) after heat stress. $(*)$ represents statistical significance measured with control and $(* *)$ represents statistical significance measured with control and between HSPB1$^{\triangle \text{C-term}}$ and HSPB1$^{-/-}$ fibroblast at P $\leq$ 0.01.

ID	Name	E-value	Region Center	Region Matches
MA0516.1	SP2	2.1E-06	107	32
MA0079.3	SP1	1.1E-05	92.5	34
MA0039.2	Klf4	1.5E-03	131	23
MA0599.1	Klf5	2.3E-02	122	25

Table 3.4: Enriched motif in down-regulated genes of samples without heat treatment (P0-C0) and E-value $\leq$ 1.

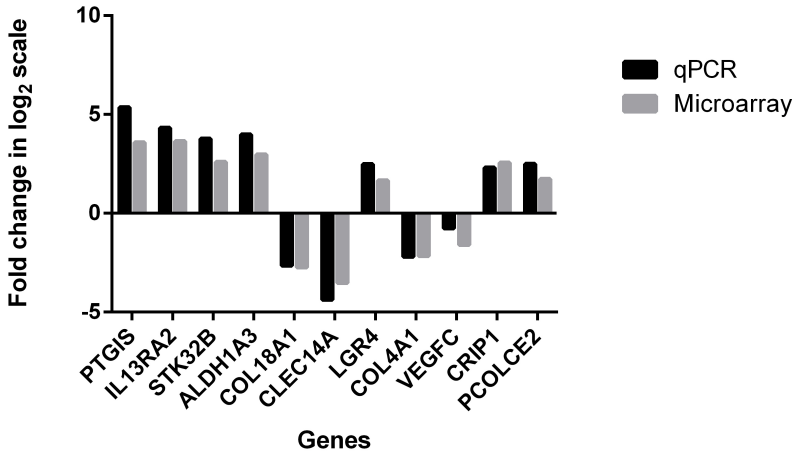


Figure 3.7: Validation of genes on altered pathway using the control (C1) as in the microarray experiment. The expression changes measured using qPCR were consistent for all the genes measured.

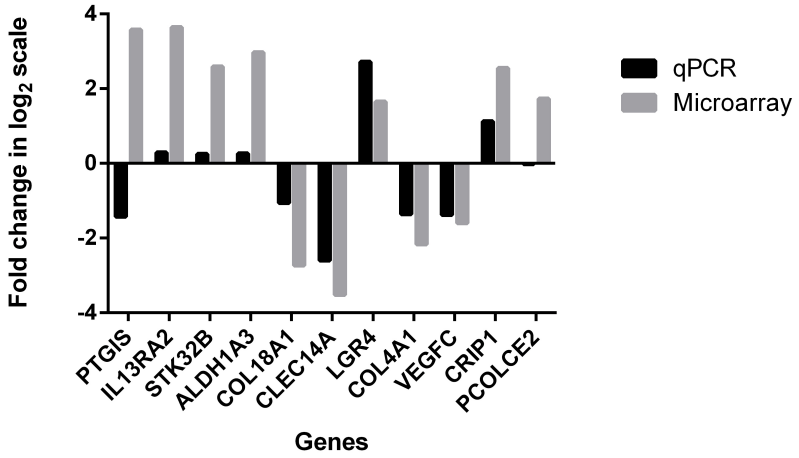


Figure 3.8: Validation of genes on altered pathway using a control (C2). The expression changes measured using qPCR were not consistent with the microarray results for all the genes measured.

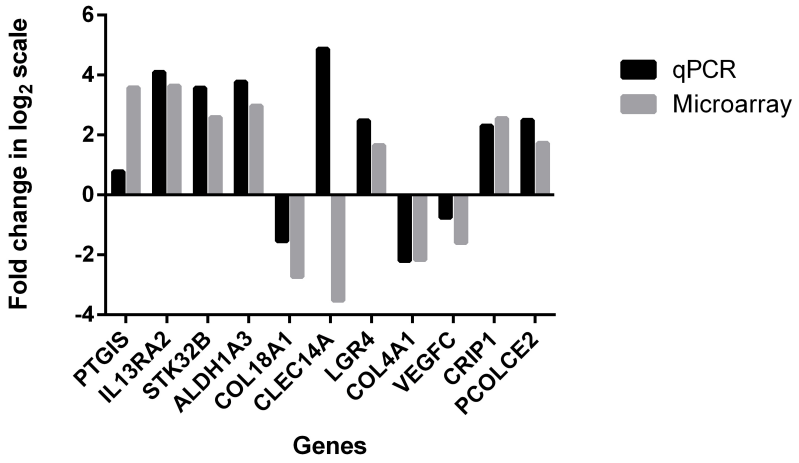


Figure 3.9: Validation of genes on altered pathway using a control (C3). The expression changes measured using qPCR were not consistent with the microarray results for all the genes measured.

3.8 qPCR Validation of genes on altered pathways

Since the microarray experiment was performed with one patient and one control sample, the results of pathway analysis and promoter analysis cannot be generalized without considering the inherent genetic variations across cultured cells. Thus, 11 genes were randomly selected from top up- and down-regulated gene list of microarray experiment belonging to different pathways including extracellular matrix proteins. The expression levels of these 11 genes were validated using three different control fibroblasts, one was the same used in the microarray experiment. Their relative expression levels were quantified using *GAPDH* as reference gene, as it had a stable expression level across different control fibroblasts. The expression levels quantified using RT-qPCR were consistent with the microarray results when observed with the same control used in the microarray experiment. However, the expression levels of selected 11 genes were inconsistent with respect to other two control fibroblasts. This could indicate that differentially expressed genes observed between patient and control cells are not only the downstream effect of truncated *HSPB1* but they could also be caused by other genetic differences between cultured cells. The expression quantified with respect to three separate controls are shown in the Figures 3.7, 3.8 and 3.9.

Chapter 4

Discussion

HSPB1 is a ubiquitously expressed multifunctional molecular chaperone that is known to be involved in a plethora of cellular processes [60]. Although being ubiquitously expressed, mutations in *HSPB1* have been associated with diseases of peripheral nervous system such as CMT2F and dHMN [60]. For example, mutations in *HSPB1* are associated with increased binding affinity to their client proteins. These increased bindings of CMT-causing *HSPB1* mutant to tubulin and microtubules result in reduction of microtubule dynamics [4]. Furthermore, mechanism of action of *HSPB1* and other sHSPs attracts attention because of their roles in protein quality control and physiologic processes [10, 41, 58]. For many diseases like neurodegeneration and cancer, understanding cell survival mechanism in stress is important because cancer cells can escape drug induced stress by regulating *HSPB1* [31]. Similarly, *HSPB1* overexpression has been found to prevent Parkinson's disease related protein aggregation [38, 58]. This study focused on characterization of the molecular consequences of a new neuropathy related *HSPB1* variant in primary patient fibroblasts. The mutation resulting in truncated *HSPB1* provided an opportunity to study the poorly understood function of C-terminus of *HSPB1* [58]. Several reported missense mutations in *HSPB1* have been studied through overexpression systems in immortalized cell lines and neuronal cells of rodents [1, 3, 4, 53]. In this study, we used primary patient's fibroblast, which adds to understanding in vivo consequences of mutant protein on endogenous level of *HSPB1* expression. We were able to show the stability of truncated *HSPB1* suggesting C.505delA mRNA escaped non-sense mediated decay leading to HSPB1$^{\triangle \text{C-term}}$ protein. This scenario is possible if premature stop codon appears in the last exon of a gene [17].

As discussed in Section 1.6, *HSPB1* can form oligomers of up to 800 kDa under normal conditions which are regulated by post translational modifi-

cations (PTMs) [41]. Dimerization is assumed to be the minimal structural unit of *HSPB1* oligomers and occurs through ACD [22, 24, 58]. In the western blot Figure 3.2, we were not able to detect any bands above 50 kDa suggesting prevalence of *HSPB1* dimers under normal conditions. Furthermore, HSPB1$^{\triangle \text{C-term}}$ fibroblasts showed bands corresponding to all possible combinations of *HSPB1* dimers. With this result, we can conclude that the HSPB1$^{\triangle \text{C-term}}$ bound to wild type protein can incur a dominant negative effect, which in turn contributed to neuropathy phenotype through reduced availability of normal wild type dimers in our patient. In a study of C-terminal variant of HSPB1$^{\text{P182L}}$, the mutant proteins were also found to form dimers with wild type *HSPB1*. Overexpressed HSPB1$^{\text{P182L}}$ in some neurons formed insoluble aggregates [1], which were not present in Figure 3.3 A of our patient fibroblasts suggesting a different pathogenic mechanism in between truncated and missense variants of C-terminus [58]. Furthermore, the aggregation of HSPB1$^{\text{P182L}}$ mutant in neurons could also be because of overexpression.

In the Chapter 3, I described nuclear translocation of *HSPB1* at high temperature. However, the nuclear role of *HSPB1* is unclear. I was able to demonstrate the translocation of HSPB1$^{\triangle \text{C-term}}$ protein to the nucleus upon heat stress, suggesting the decreased heat tolerance of patient cells is not caused by defective translocation [58]. Moreover, expression of similar genes in both treated (P30-C30) and untreated (P0-C0) condition of the microarray experiment and enrichment of pathways like macrophage activation, cellular defence response, response to stimulus, immune system processes are all expected as patient fibroblast with truncated *HSPB1* has reduced ability to respond against stress. Furthermore, heat itself has a denaturing effect on proteins causing additional stress. Thus, increased expression levels of genes involved in these pathways provide ideas on how cells are trying to compensate the stress. These results suggest the C-terminal truncation of *HSPB1* did not influence the early heat shock response. However, enrichment of extracellular matrix proteins was unexpected under both treated and untreated conditions. This could mean that *HSPB1* has an extracellular role or participates in signal transduction. Although the extracellular role of *HSPB1* is unclear, few articles have corroborated the fact that the secreted *HSPB1* participates in chaperoning immunogenic peptides and they were able to detect extracellular *HSPB1* through western blot [33, 47]. Following the protocol of [33], I was not able to detect a band corresponding to wild type or truncated *HSPB1* from the conditioned medium. This might be because of the low concentration of protein in the medium. Furthermore, the secretion of *HSPB1* might also be cell type specific as previous findings were derived from acti-

vated macrophages and endothelial cells. There are other sensitive methods to detect secreted proteins from medium such as enzyme-linked immunosorbent assay (ELISA), mass spectroscopy (MS), but these analysis were out of scope for this thesis.

Moreover, RT-qPCR validation of genes on altered pathways revealed some inconsistencies. As stated before, the microarray experiment was performed with one patient and one control sample. Although the expression levels of selected 11 genes measured by RT-qPCR for the same control were consistent with the microarray expression levels, the measurements for the same 11 genes for other two controls did not agree with earlier results. This suggests the changes in gene expression of patient cells were not entirely because of the truncated *HSPB1* but also other genetic differences between the cultured cells, which points to the need for more reliable and genetically homogeneous control sample, which would reduce the genetic background. One method to achieve this is to correct the mutation (C.505delA) using CRISPR/Cas9 in patient fibroblasts and use it as a control sample.

HSPB1 response to external stresses like heat, oxidative stress, heavy metals and ischemia protect cells and also is a general theme of protection of sHSPs [58]. Heat is a generalized stressor and it causes protein misfolding, loss of host cell functions, induces cytoskeleton and cell cycle arrest [56, 58]. sHSPs differ from normal heat shock proteins, mainly because these sHSPs lack ATPase activity, hence cannot actively refold proteins [41]. HSPB1 binds to denatured or misfolded proteins and prevents the formation of toxic aggregates [10]. Using the HSPB1$^{\triangle\text{C-term}}$, HSPB1$^{-/-}$ and control fibroblasts we were able to demonstrate significantly decreased heat tolerance in both HSPB1$^{\triangle\text{C-term}}$ and HSPB1$^{-/-}$ fibroblasts in comparison to control cells, but growth of HSPB1$^{\triangle\text{C-term}}$ and HSPB1$^{-/-}$ did not differ significantly until the last time point (refer Figure 3.6). With a longer recovery we might be able to see a difference but this result suggests reduced dimer formation by a dominant-negative effect has the same consequence as having no HSPB1 at all.

Similarly, response of the patient (P0-P30) and response of the control (C0-C30) to heat shock yielded 5 differentially expressed genes under both conditions. The genes were coding for proteins that protects cells from stress (*DNAJB1, EGR1, FOS, HSPA6* and *HSPA7*). This result suggests patient and control fibroblasts have similar transcriptional heat shock response. Furthermore, there were small changes in the expression of the other 45 genes that were statistically significant for response of control to heat shock. Al-

though these small changes in expression might have large effect in post mitotic cells, these changes usually have little or no significance in actively dividing cells like fibroblast. Thus, we cannot conclusively assert the fact that the transcriptional response of patient fibroblast is different than that of control fibroblast. To confirm these findings, it would be interesting to observe the gene expression profiles in neurons derived from the patient and control fibroblasts before and after heat stress. In addition to microarray results, expression levels quantified using qRT-PCR for two of the differentially expressed heat shock proteins (*DNAJB1* and *HSPA6*) in control, patient and HSPB1 knockout (HSPB1$^{-/-}$) fibroblasts suggest that *HSPB1* might not have any role in the expression of those genes. This is the indication that C-terminal truncation did not participate in early induction of heat-stress related genes [58]. Thus, *HSPB1* may have a chaperoning function of misfolded proteins in the nucleus rather than direct involvement in regulation of gene expression. The reduction in the growth rate of HSPB1$^{\triangle \text{C-term}}$ fibroblasts during extended exposure to heat, might have been caused by denaturation of proteins responsible for cell growth and cell division, which the truncated *HSPB1* was unable to chaperone [58]. With these results it can be stated that the C-terminus of *HSPB1* is necessary for the protein to maintain its ability to protect cell against partially denatured, independent of nuclear translocation [58]. Since the results of microarray experiment and pathway analysis is based on one patient and one control sample, the statistical power and biological relevance of the analysis is reduced as an inherent genetic variability of cultured cells remains unaccounted.

Analysis of the promoter region of down-regulated genes of patient fibroblasts revealed enrichment of binding sites for transcription factors *SP1, SP2, KLF4* and *KLF5*. Studies have identified *SP1* transcription factor binding sites at promoter regions of growth related genes providing evidence for its involvement in cell growth. This function of *SP1* was confirmed when truncated *SP1* was able to inhibit growth of cells through a dominant negative mechanism [13]. SP/KLF is family of transcription factors of 20 identified members with a characteristic zinc finger domain similar to *Drosophila* protein, *krüppel* [13]. SP/KLF factors bind to GC-boxes in DNA with varying affinities and potentially affect "SP1 dependent" transcription [13]. Regulation of members of the SP/KLF family in the developmental phase corroborates its involvement in cell growth and differentiation [13]. Enrichment of these transcription factors in the promoters of differentially expressed down-regulated genes suggest that, the decreased growth rate of HSPB1$^{\triangle \text{C-term}}$ fibroblasts might have been caused by reduced ability of truncated *HSPB1* to chaperone these misfolded transcription factors during heat shock or re-

duced availability of wild type *HSPB1* dimer by a dominant negative effect. Furthermore, molecular mechanism of KLF transcription factors has been studied in proliferating cells, but limited is known in post mitotic cells like neurons where even small changes in steady state concentration of these proteins can have a large effect. One of the noted effects is decreased neurite outgrowth by expression of *KLF4* in hippocampal and cortical neurons [39]. Although our expression results are based on proliferating patient fibroblast cells, it would be interesting to study the expression levels of KLF family in iPSc (induced pluripotent stem cells) derived patient neurons.

Chapter 5

Limitations of the study and future work

The major limitation of this study is the small sample size, as only one HSPB1$^{\triangle \text{C-term}}$ patient and one HSPB1$^{-/-}$ fibroblasts were available. The microarray experiment was based on only one patient and control sample and that has reduced the statistical power of the analysis in the project. A genetically homogeneous control sample would have reduced the background expression and would have provided better insight on altered pathway and functionality of C-terminus of HSPB1. Another microarray experiment with control generated through correction of the truncated mutation using CRISPR in patient fibroblast would reduce the genetic background and provide more accurate findings. Furthermore, we were able to show that HSPB1$^{\triangle \text{C-term}}$ is associated with peripheral neuropathy and impairs heat tolerance of cells, but more accurate disease models such as iPS cells derived patient neurons are needed to fully understand the pathogenesis of HSPB1 related neuropathy [58]. Moreover, more patient samples with same or similar mutations would allow more extensive testing of the conclusions derived from this study.

It would be interesting to investigate the stress tolerance in neuronal cells and differentiated neurons from patient-derived induced pluripotent stem cells [58]. Since, peripheral motor and sensory neurons are post-mitotic they are sensitive to small changes in steady state concentration of cellular components. Thus, it would also be interesting to test the findings of promoter analysis and check the expression level of KLF transcription factor family specially KLF4 and KLF5 in differentiated neurons from patient-derived induced pluripotent stem cells.

Bibliography

[1] Ackerley, S., James, P. A., Kalli, A., French, S., Davies, K. E., and Talbot, K. A mutation in the small heat-shock protein hspb1 leading to distal hereditary motor neuronopathy disrupts neurofilament assembly and the axonal transport of specific cellular cargoes. *Human molecular genetics 15*, 2 (2006), 347–354.

[2] Ajroud-Driss, S., Fecto, F., Ajroud, K., Lalani, I., Calvo, S. E., Mootha, V. K., Deng, H.-X., Siddique, N., Tahmoush, A. J., Heiman-Patterson, T. D., et al. Mutation in the novel nuclear-encoded mitochondrial protein chchd10 in a family with autosomal dominant mitochondrial myopathy. *neurogenetics 16*, 1 (2015), 1–9.

[3] Almeida-Souza, L., Asselbergh, B., d'Ydewalle, C., Moonens, K., Goethals, S., de Winter, V., Azmi, A., Irobi, J., Timmermans, J.-P., Gevaert, K., et al. Small heat-shock protein hspb1 mutants stabilize microtubules in charcot-marie-tooth neuropathy. *The Journal of Neuroscience 31*, 43 (2011), 15320–15328.

[4] Almeida-Souza, L., Goethals, S., de Winter, V., Dierick, I., Gallardo, R., Van Durme, J., Irobi, J., Gettemans, J., Rousseau, F., Schymkowitz, J., et al. Increased monomerization of mutant hspb1 leads to protein hyperactivity in charcot-marie-tooth neuropathy. *Journal of biological chemistry 285*, 17 (2010), 12778–12786.

[5] Auranen, M., Ylikallio, E., Shcherbii, M., Paetau, A., Kiuru-Enari, S., Toppila, J. P., and Tyynismaa, H. Chchd10 variant p.(gly66val) causes axonal charcot-marie-tooth disease. *Neurology Genetics 1*, 1 (2015), e1.

[6] Babu, M. M. Introduction to microarray data analysis. *Computational Genomics: Theory and Application* (2004), 225–249.

[7] Bailey, T. L., and Machanick, P. Inferring direct dna binding from chip-seq. *Nucleic acids research* (2012), gks433.

[8] Bannwarth, S., Ait-El-Mkadem, S., Chaussenot, A., Genin, E. C., Lacas-Gervais, S., Fragaki, K., Berg-Alonso, L., Kageyama, Y., Serre, V., Moore, D. G., et al. A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through chchd10 involvement. *Brain 137*, 8 (2014), 2329–2345.

[9] Baranova, E., Weeks, S., Beelen, S., Bukach, O., Gusev, N., and Strelkov, S. Three-dimensional structure of α-crystallin domain dimers of human small heat shock proteins hspb1 and hspb6. *Journal of molecular biology 411*, 1 (2011), 110–122.

[10] Basha, E., O'Neill, H., and Vierling, E. Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions. *Trends in biochemical sciences 37*, 3 (2012), 106–117.

[11] Bellofatto, V., and Wilusz, J. Transcription and mrna stability: parental guidance suggested. *Cell 147*, 7 (2011), 1438–1439.

[12] Benndorf, R., Martin, J. L., Pond, S. L. K., and Wertheim, J. O. Neuropathy-and myopathy-associated mutations in human small heat shock proteins: Characteristics and evolutionary history of the mutation sites. *Mutation Research/Reviews in Mutation Research 761* (2014), 15–30.

[13] Black, A. R., Black, J. D., and Azizkhan-Clifford, J. Sp1 and krüppel-like factor family of transcription factors in cell growth regulation and cancer. *Journal of cellular physiology 188*, 2 (2001), 143–160.

[14] Bryantsev, A. L., Loktionova, S. A., Ilyinskaya, O. P., Tararak, E. M., Kampinga, H. H., and Kabakov, A. E. Distribution, phosphorylation, and activities of hsp25 in heat-stressed h9c2 myoblasts: a functional link to cytoprotection. *Cell stress & chaperones 7*, 2 (2002), 146.

[15] Butte, A. The use and analysis of microarray data. *Nature reviews drug discovery 1*, 12 (2002), 951–960.

[16] Chalova, A. S., Sudnitsyna, M. V., Strelkov, S. V., and Gusev, N. B. Characterization of human small heat shock protein hspb1

that carries c-terminal domain mutations associated with hereditary motor neuron diseases. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics 1844*, 12 (2014), 2116–2126.

[17] Chang, Y.-F., Imam, J. S., and Wilkinson, M. F. The nonsense-mediated decay rna surveillance pathway. *Annu. Rev. Biochem. 76* (2007), 51–74.

[18] Chen, X., Guo, L., Fan, Z., and Jiang, T. Learning position weight matrices from sequence and expression data. In *Comput Syst Bioinformatics Conf* (2007), vol. 6, World Scientific, pp. 249–260.

[19] Czeizler, E. High-throughput bioinformatics: Microarrays part ii. University Lecture, 2014.

[20] Dreos, R., Ambrosini, G., Périer, R. C., and Bucher, P. Epd and epdnew, high-quality promoter resources in the next-generation sequencing era. *Nucleic acids research 41*, D1 (2013), D157–D164.

[21] Du, P., Kibbe, W.A., Lin, and S.M. lumi: a pipeline for processing illumina microarray. *Bioinformatics* (2008).

[22] Dudich, I. V., Zav'yalov, V. P., Pfeil, W., Gaestel, M., Zav'yalova, G. A., Denesyuk, A. I., and Korpela, T. Dimer structure as a minimum cooperative subunit of small heat-shock proteins. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology 1253*, 2 (1995), 163–168.

[23] d'Ydewalle, C., Krishnan, J., Chiheb, D. M., Van Damme, P., Irobi, J., Kozikowski, A. P., Berghe, P. V., Timmerman, V., Robberecht, W., and Van Den Bosch, L. Hdac6 inhibitors reverse axonal loss in a mouse model of mutant hspb1-induced charcot-marie-tooth disease. *Nature medicine 17*, 8 (2011), 968–974.

[24] Ehrnsperger, M., Lilie, H., Gaestel, M., and Buchner, J. The dynamics of hsp25 quaternary structure structure and function of different oligomeric species. *Journal of Biological Chemistry 274*, 21 (1999), 14867–14874.

[25] Evgrafov, O. V., Mersiyanova, I., Irobi, J., Van Den Bosch, L., Dierick, I., Leung, C. L., Schagina, O., Verpoorten, N., Van Impe, K., Fedotov, V., et al. Mutant small heat-shock protein 27 causes axonal charcot-marie-tooth disease and distal hereditary motor neuropathy. *Nature genetics 36*, 6 (2004), 602–606.

[26] Friedberg, E. C., Walker, G. C., Siede, W., and Wood, R. D. *DNA repair and mutagenesis*. American Society for Microbiology Press, 2005.

[27] Guigo, R. An introduction to position specific scoring matrices, Apr. 2003. `http://bioinformatica.upf.edu/T12/MakeProfile.html`.

[28] Hochberg, G. K., Ecroyd, H., Liu, C., Cox, D., Cascio, D., Sawaya, M. R., Collier, M. P., Stroud, J., Carver, J. A., Baldwin, A. J., et al. The structured core domain of αb-crystallin can prevent amyloid fibrillation and associated toxicity. *Proceedings of the National Academy of Sciences 111*, 16 (2014), E1562–E1570.

[29] Houlden, H., Laura, M., Wavrant-De Vrièze, F., Blake, J., Wood, N., and Reilly, M. Mutations in the hsp27 (hspb1) gene cause dominant, recessive, and sporadic distal hmn/cmt type 2. *Neurology 71*, 21 (2008), 1660–1668.

[30] Irobi, J., Van Impe, K., Seeman, P., Jordanova, A., Dierick, I., Verpoorten, N., Michalik, A., De Vriendt, E., Jacobs, A., Van Gerwen, V., et al. Hot-spot residue in small heat-shock protein 22 causes distal motor neuropathy. *Nature genetics 36*, 6 (2004), 597–601.

[31] Katsogiannou, M., Andrieu, C., and Rocchi, P. Heat shock protein 27 phosphorylation state is associated with cancer progression. *Frontiers in genetics 5* (2014).

[32] Kolb, S., Snyder, P., Poi, E., Renard, E., Bartlett, A., Gu, S., Sutton, S., Arnold, W., Freimer, M., Lawson, V., et al. Mutant small heat shock protein b3 causes motor neuropathy utility of a candidate gene approach. *Neurology 74*, 6 (2010), 502–506.

[33] Lee, Y.-J., Lee, H.-J., Choi, S.-h., Jin, Y. B., An, H. J., Kang, J.-H., Yoon, S. S., and Lee, Y.-S. Soluble hspb1 regulates vegf-mediated angiogenesis through their direct interaction. *Angiogenesis 15*, 2 (2012), 229–242.

[34] Lupski, J. R., Reid, J. G., Gonzaga-Jauregui, C., Rio Deiros, D., Chen, D. C., Nazareth, L., Bainbridge, M., Dinh, H., Jing, C., Wheeler, D. A., et al. Whole-genome sequencing in a patient with charcot–marie–tooth neuropathy. *New England Journal of Medicine 362*, 13 (2010), 1181–1191.

[35] Mandich, P., Grandis, M., Varese, A., Geroldi, A., Acquaviva, M., Ciotti, P., Gulli, R., Doria-Lamba, L., Fabrizi, G. M., Giribaldi, G., et al. Severe neuropathy after diphtheria-tetanus-pertussis vaccination in a child carrying a novel frame-shift mutation in the small heat-shock protein 27 gene. *Journal of child neurology 25*, 1 (2010), 107–109.

[36] McLeay, R. C., and Bailey, T. L. Motif enrichment analysis: a unified framework and an evaluation on chip data. *BMC bioinformatics 11*, 1 (2010), 165.

[37] Mi, H., Muruganujan, A., Casagrande, J. T., and Thomas, P. D. Large-scale gene function analysis with the panther classification system. *Nature protocols 8*, 8 (2013), 1551–1566.

[38] Minoia, M., Grit, C., and Kampinga, H. H. Hspa1a-independent suppression of park2 c289g protein aggregation by human small heat shock proteins. *Molecular and cellular biology 34*, 19 (2014), 3570–3578.

[39] Moore, D. L., Blackmore, M. G., Hu, Y., Kaestner, K. H., Bixby, J. L., Lemmon, V. P., and Goldberg, J. L. Klf family members regulate intrinsic axon regeneration ability. *Science 326*, 5950 (2009), 298–301.

[40] Moss, A. C., Doran, P. P., and MacMathuna, P. In silico promoter analysis can predict genes of functional relevance in cell proliferation: validation in a colon cancer model. *Translational oncogenomics 2* (2007), 1.

[41] Mymrikov, E. V., Seit-Nebi, A. S., and Gusev, N. B. Large potentials of small heat shock proteins. *Physiological reviews 91*, 4 (2011), 1123–1159.

[42] NIH. Charcot-marie-tooth disease, June 2013. `http://www.ninds.nih.gov/disorders/charcot_marie_tooth/charcot-marie-tooth_brochure_508comp.pdf`.

[43] NIH. What are the different ways in which a genetic condition can be inherited?, July 2015. `http://ghr.nlm.nih.gov/handbook/inheritance/inheritancepatterns`.

[44] Nishida, K., Frith, M. C., and Nakai, K. Pseudocounts for transcription factor binding sites. *Nucleic acids research 37*, 3 (2009), 939–944.

[45] Pereyra, N. B. *In silico analysis of regulatory motifs in gene promoters*. Universitat Pompeu Fabra, 2010.

[46] Pranovo. Western blot transfer, Nov. 2012. `https://novowb.wordpress.com/2012/11/10/western-blot-transfer/`.

[47] Rayner, K., Chen, Y.-X., McNulty, M., Simard, T., Zhao, X., Wells, D. J., de Belleroche, J., and O'Brien, E. R. Extracellular release of the atheroprotective heat shock protein 27 is mediated by estrogen and competitively inhibits acldl binding to scavenger receptor-a. *Circulation research 103*, 2 (2008), 133–141.

[48] Reilly, M. M., Murphy, S. M., and Laurá, M. Charcot-marie-tooth disease. *Journal of the Peripheral Nervous System 16*, 1 (2011), 1–14.

[49] Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research 43* (2015), doi: 10.1093/nar/gkv007.

[50] Rossor, A. M., Davidson, G. L., Blake, J., Polke, J. M., Murphy, S. M., Houlden, H., Innes, A., Kalmar, B., Greensmith, L., and Reilly, M. M. A novel p. glu175x premature stop mutation in the c-terminal end of hsp27 is a cause of cmt2. *Journal of the Peripheral Nervous System 17*, 2 (2012), 201–205.

[51] Rossor, A. M., Polke, J. M., Houlden, H., and Reilly, M. M. Clinical implications of genetic advances in charcot–marie–tooth disease. *Nature Reviews Neurology 9*, 10 (2013), 562–571.

[52] Schneider, T. D. Consensus sequence zen. *Applied bioinformatics 1*, 3 (2002), 111.

[53] Selcen, D., and Engel, A. G. Myofibrillar myopathy caused by novel dominant negative αb-crystallin mutations. *Annals of neurology 54*, 6 (2003), 804–810.

[54] Stromer, T., Ehrnsperger, M., Gaestel, M., and Buchner, J. Analysis of the interaction of small heat shock proteins with unfolding proteins. *Journal of Biological Chemistry 278*, 20 (2003), 18015–18021.

[55] Szigeti, K., and Lupski, J. R. Charcot–marie–tooth disease. *European Journal of Human Genetics 17*, 6 (2009), 703–710.

[56] Velichko, A. K., Markova, E. N., Petrova, N. V., Razin, S. V., and Kantidze, O. L. Mechanisms of heat shock response in mammals. *Cellular and Molecular Life Sciences 70*, 22 (2013), 4229–4241.

[57] Vicart, P., Caron, A., Guicheney, P., Li, Z., Prévost, M.-C., Faure, A., Chateau, D., Chapon, F., Tomé, F., Dupret, J.-M., et al. A missense mutation in the &agr; b-crystallin chaperone gene causes a desmin-related myopathy. *Nature genetics 20*, 1 (1998), 92–95.

[58] Ylikallio, E., Konovalova, S., Dhungana, Y., Hilander, T., Junna, N., Partanen, J. V., Toppila, J. P., Auranen, M., and Tyynismaa, H. Truncated hspb1 causes axonal neuropathy and impairs tolerance to unfolded protein stress. *BBA Clinical 3*, 0 (2015), 233 – 242.

[59] Zhai, J., Lin, H., Julien, J.-P., and Schlaepfer, W. W. Disruption of neurofilament network with aggregation of light neurofilament protein: a common pathway leading to motor neuron degeneration due to charcot–marie–tooth disease-linked mutations in nfl and hspb1. *Human molecular genetics 16*, 24 (2007), 3103–3116.

[60] Züchner, S., and Vance, J. M. Mechanisms of disease: a molecular genetic update on hereditary axonal neuropathies. *Nature Clinical Practice Neurology 2*, 1 (2006), 45–53.

Appendix A

Appendix

Gene	Primer Sequence (5'→ 3')
IL13RA2_F	CAATGCTGGGAAGGTGAAGAC
IL13RA2_R	TGGGTAGGTGTTTGGCTTACG
ALDH1A3_F	AGGGTGGGCAGACAAAATCC
ALDH1A3_R	GGGGAAGTTCCATGGAGTGAT
STK32B_F	GAATATGGGCGGGAACCACT
STK32B_R	TGCACGATGCATACCTTTCC
COL18A1_F	GCCGTGGCATTCCTAGCTC
COL18A1_R	CTGATGCGCTCTGAAGATGGT
CLEC14A_F	GGGATCACAGAGCACGATGT
CLEC14A_R	CCTGATGGGGTGATAGTGGC
PTGIS_F	CTGGAGGAGATGGGTGTGTC
PTGIS_R	GGACCCATATTCCCCTGTGT
LGR4_F	GAAGAGCTACAATTGGCGGG
LGR4_R	CTCGAATGGCTTCACTGGGT
COL4A1_F	ACTCTTTTGTGATGCACACCA
COL4A1_R	AAGCTGTAAGCGTTTGCGTA
VEGFC_F	GAGGCTGGCAACATAACAGAG
VEGFC_R	CCTTGAGAGAGAGGCACTGT
CRIP1_F	AGTAAACCAGGTGGTGGAGAC
CRIP1_R	AGCATTAGGGGCAACAAGGG
PCOLCE2_F	TAATGGCGGGGAAGTCAACG
PCOLCE2_R	AATTGGCGCAGGTGGACTAT
HSPA6_F	CAGAGGAACGCCACTATCCC
HSPA6_R	ACTGAGTTCAAAACGCCCCA
DNAJB1_F	CTGCGCTACCACCCGGAC
DNAJB1_R	TTTAGGCCTTCCTCCCCGTAG

Table A.1: Primers designed for the project.

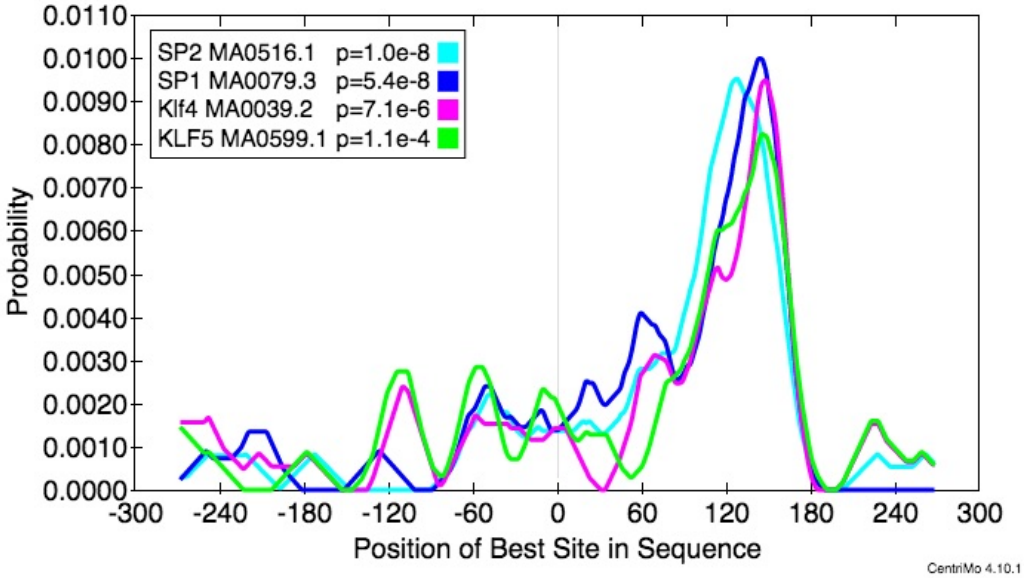


Figure A.1: Predicted occupancy site of enriched transcription factor in the promoter region of down-regulated genes from samples without heat treatment (P0-C0).